

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**

318 Nephrolithiasis	2368
<i>Gary C. Curhan</i>	
319 Urinary Tract Obstruction	2373
<i>Julian L. Seifert</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>Haralampus 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>Haralampus M. Moutsopoulos, Clio P. Mavragani</i>	
362 Spondyloarthritis	2790
<i>Joel D. Taurog, Lianne S. Gensler, Nigel Haroon</i>	
363 The Vasculitis Syndromes	2802
<i>Carol A. Langford, Anthony S. Fauci</i>	
364 Behçet Syndrome.....	2817

Section 1 Disorders of the Alimentary Tract

321

Approach to the Patient with Gastrointestinal Disease

William L. Hasler, Chung Owyang



ANATOMIC CONSIDERATIONS

The gastrointestinal (GI) tract extends from the mouth to the anus and is composed of organs with distinct functions. Sphincters that assist in gut compartmentalization separate the organs. The gut wall is organized into distinct layers that contribute to regional activities. The mucosa is a barrier to luminal contents or a site for fluid and nutrient transfer. Smooth muscle in association with the enteric nervous system mediates propulsion between regions. Many GI organs possess a serosal layer that provides a supportive foundation and permits external input.

Interactions with other systems serve the needs of the gut and the body. Pancreaticobiliary conduits deliver bile and enzymes into the duodenum. The vascular supply is modulated by GI activity. Lymphatic channels assist in gut immune activities. Intrinsic nerves provide the controls for propulsion and fluid regulation. Extrinsic neural input provides volitional or involuntary control that is specific for each gut region.

FUNCTIONS OF THE GI TRACT

The GI tract serves two main functions—assimilating nutrients and eliminating waste. In the mouth, food is processed, mixed with salivary amylase, and delivered to the gut lumen. The esophagus propels the bolus into the stomach; the lower esophageal sphincter prevents oral reflux of gastric contents. The squamous esophageal mucosa protects against significant diffusion or absorption. Aboral esophageal contractions coordinate with relaxation of the upper and lower esophageal sphincters on swallowing.

The stomach triturates and mixes the food bolus with pepsin and acid. Gastric acid also sterilizes the upper gut. The proximal stomach serves a storage function by relaxing to accommodate the meal. Phasic contractions in the distal stomach propel food residue against the pylorus, where it is ground and thrust proximally for further mixing before it is emptied into the duodenum. The stomach secretes intrinsic factor for vitamin B₁₂ absorption.

Most nutrient absorption occurs in the small intestine. The mucosal villus architecture provides maximal surface area for absorption and is endowed with specialized enzymes and transporters. Triturated food from the stomach mixes with pancreatic juice and bile in the duodenum. Pancreatic juice contains enzymes for nutrient digestion and bicarbonate to optimize the pH for enzyme activation. Bile secreted by the liver and stored in the gallbladder is essential for lipid digestion. The proximal intestine is optimized for rapid absorption of most nutrients and minerals, whereas the ileum is better suited for absorbing vitamin B₁₂ and bile acids. Bile contains by-products of erythrocyte degradation, toxins, medications, and cholesterol for fecal evacuation. Intestinal motor function delivers indigestible residue into the colon for processing. The ileocecal junction is a sphincter that prevents coloileal reflux, reducing microbial density.

The colon prepares waste for evacuation. The mucosa dehydrates the stool, reducing daily ileal volumes of 1000–1500 mL to 100–200 mL

expelled from the rectum. The colon possesses a dense bacterial colonization that ferments undigested carbohydrates and short-chain fatty acids. The gut microbiome also modulates immune and physiologic activity. Esophageal transit takes seconds, and times in the stomach and small intestine range from minutes to a few hours, but colon propagation requires >1 day in most individuals. Colon contractions exhibit a to-and-fro character that promotes fecal desiccation. The proximal colon mixes and absorbs fluid, while the distal colon exhibits peristaltic contractions and mass movements to expel the stool. The colon terminates in the anus, which possesses volitional and involuntary controls to permit fecal retention until it can be released in a convenient setting.

EXTRINSIC MODULATION OF GUT FUNCTION

GI function is modified by influences outside the gut. Unlike other organs, the gut is in continuity with the outside environment. Protective mechanisms are vigilant against injury from foods, medications, toxins, and microbes. Mucosal immune mechanisms include epithelial and lamina propria lymphocytes and plasma cells supported by lymph node chains to prevent noxious agents from entering the circulation. Antimicrobial peptides secreted by Paneth cells defend against pathogens. Drugs and toxins absorbed into the bloodstream are filtered and detoxified in the liver via the portal venous circulation. Although intrinsic nerves control most basic gut activities, extrinsic neural input modulates many functions. Many GI reflexes involve extrinsic vagus or splanchnic nerve pathways. The brain-gut axis alters function in regions not under volitional regulation. Stress can disrupt gut motor, secretory, and sensory function.

OVERVIEW OF GI DISEASES

GI diseases develop as a result of abnormalities within or outside of the gut and range in severity from those that produce mild symptoms and no long-term morbidity to those with intractable symptoms or adverse outcomes. Diseases may be localized to one organ or exhibit diffuse involvement at many sites.

CLASSIFICATION OF GI DISEASES

GI diseases are manifestations of alterations in nutrient assimilation or waste evacuation or in the activities supporting these main functions.

Impaired Digestion and Absorption Diseases of the stomach, intestine, biliary tree, and pancreas can disrupt digestion and absorption. The most common maldigestion syndrome, lactase deficiency, produces gas and diarrhea after ingesting dairy products and has no adverse outcomes. Other intestinal enzyme deficiencies produce similar symptoms after consuming other simple sugars. Celiac disease, bacterial overgrowth, infectious enteritis, Crohn's ileitis, and radiation damage, which affect digestion and/or absorption more diffusely, produce anemia, dehydration, electrolyte disorders, or malnutrition. Gastric hypersecretory conditions such as gastrinoma damage the intestinal mucosa, impair pancreatic enzyme activation, and accelerate transit due to excess gastric acid. Benign or neoplastic biliary obstruction impairs fat digestion. Impaired pancreatic enzyme release in chronic pancreatitis or pancreatic cancer decreases intraluminal digestion and can lead to malnutrition.

Altered Secretion Some GI diseases result from dysregulation of gut secretion. Gastric acid hypersecretion occurs in gastrinoma, G-cell hyperplasia, retained antrum syndrome, and some patients with duodenal ulcers. Gastric acid is reduced in atrophic gastritis and pernicious anemia. Inflammatory and infectious small-intestinal and colonic diseases produce fluid loss through impaired absorption or enhanced secretion. Common hypersecretory conditions that cause diarrhea include acute bacterial or viral infection, chronic *Giardia* or cryptosporidium infections, small-intestinal bacterial overgrowth, bile salt diarrhea, microscopic colitis, and diabetic diarrhea. Less common

causes include large colonic villus adenomas and endocrine neoplasias with tumor overproduction of secretagogue transmitters such as vasoactive intestinal polypeptide.

Altered Gut Transit Impaired gut transit may result from mechanical obstruction. Esophageal occlusion most often is due to stricture (due to acid exposure or eosinophilic esophagitis) or neoplasm. Gastric obstruction develops from ulcer disease or gastric cancer. Small-intestinal obstruction most commonly results from adhesions but also occurs with Crohn's disease, radiation- or drug-induced strictures, and less likely malignancy. The most common cause of colonic obstruction is colon cancer, although inflammatory strictures develop with inflammatory bowel disease (IBD), after certain infections such as diverticulitis, or with some drugs.

Retardation of propulsion can develop from altered motor function. Achalasia is characterized by impaired esophageal body peristalsis and incomplete lower esophageal sphincter relaxation. Gastroparesis is the delay in gastric emptying of meals due to impaired gastric motility. Intestinal pseudoobstruction is the disruption of small-bowel contractility due to enteric nerve or smooth-muscle injury. Slow-transit constipation results from diffusely impaired colon propulsion. Constipation also is produced by outlet abnormalities such as rectal prolapse, intussusception, or dyssynergia—a failure of anal or puborectalis relaxation upon attempted defecation.

Disorders of rapid propulsion are less common than those with delayed transit. Rapid gastric emptying occurs with postvagotomy dumping syndrome, gastric hypersecretion, and some cases of functional dyspepsia and cyclic vomiting syndrome. Exaggerated intestinal or colonic motor patterns may be responsible for diarrhea in irritable bowel syndrome (IBS). Accelerated transit with hyperdefecation is noted in hyperthyroidism.

Immune Dysregulation Many inflammatory GI conditions are consequences of altered gut immune function. Mucosal inflammation in celiac disease results from dietary ingestion of gluten-containing grains. Some patients with food allergy also exhibit altered immune populations. Eosinophilic esophagitis and eosinophilic gastroenteritis are inflammatory disorders with prominent mucosal eosinophil infiltration. Ulcerative colitis and Crohn's disease are disorders that produce mucosal injury primarily in the lower gut. The microscopic colitides, lymphocytic and collagenous colitis, exhibit colonic subepithelial infiltrates without visible mucosal damage. Bacterial, viral, and protozoal organisms produce ileitis or colitis in selected patients. Alterations in the gut microbiome (termed dysbiosis) are proposed to trigger IBD, celiac disease, and IBS flares and may be factors in oncogenesis in some cases of pancreatic cancer.

Impaired Gut Blood Flow Different GI regions are at variable risk for ischemic damage from impaired blood flow. Rare cases of gastroparesis result from blockage of the celiac and superior mesenteric arteries. More commonly encountered are intestinal and colonic ischemia that are consequences of arterial embolus, arterial thrombosis,

venous thrombosis, or hypoperfusion from dehydration, sepsis, hemorrhage, or reduced cardiac output. These may produce mucosal injury, hemorrhage, or even perforation. Chronic ischemia may result in intestinal stricture. Some cases of radiation enterocolitis exhibit reduced mucosal blood flow.

Neoplastic Degeneration All GI regions are susceptible to malignant degeneration. In the United States, colorectal cancer is most common and usually presents after age 45 years. Worldwide, gastric cancer is prevalent, especially in certain Asian populations. Esophageal cancer develops with chronic acid reflux or after extensive alcohol or tobacco use. Small-intestinal neoplasms are rare but occur with underlying inflammatory diseases. Anal cancers arise after prior anal infection or inflammation. Pancreatic and biliary cancers elicit severe pain, weight loss, and jaundice and have poor prognoses. Hepatocellular carcinoma usually arises in the setting of chronic viral hepatitis or cirrhosis secondary to other causes. Most GI cancers exhibit carcinomatous histology; however, lymphomas and other cell types also are observed.

Disorders without Obvious Organic Abnormalities The most prevalent GI disorders show no abnormalities on biochemical or structural testing and include IBS, functional dyspepsia, and functional heartburn. These disorders exhibit altered gut motor function, but the pathogenic relevance of these abnormalities is uncertain. Exaggerated visceral sensory responses to noxious stimulation may cause discomfort in these disorders. Symptoms in other patients result from altered processing of visceral pain sensations in the central nervous system. Functional bowel patients with severe symptoms may exhibit significant emotional disturbances on psychometric testing. Subtle immunologic defects may contribute to functional symptoms as well.

Genetic Influences Although many GI diseases result from environmental factors, others exhibit hereditary components. Family members of IBD patients show a genetic predisposition to disease development themselves. Colonic, esophageal, and pancreatic malignancies arise in certain inherited disorders. Rare genetic dysmotility syndromes are described. Familial clustering is observed in the functional bowel disorders, although this may be secondary learned familial illness behavior rather than a true hereditary factor.

■ SYMPTOMS OF GI DISEASE

Symptoms of GI disease include abdominal pain, heartburn, nausea and vomiting, altered bowel habits, GI bleeding, jaundice, and other manifestations (Table 321-1).

Abdominal Pain Abdominal pain results from GI disease and extraintestinal conditions involving the genitourinary tract, abdominal wall, thorax, or spine. Visceral pain generally is midline in location and vague in character, whereas parietal pain is localized and precisely described. Painful inflammatory diseases include peptic ulcer, appendicitis, diverticulitis, IBD, pancreatitis, cholecystitis, and infectious enterocolitis. Noninflammatory visceral sources include biliary colic,

TABLE 321-1 Common Causes of Common Gastrointestinal (GI) Symptoms

ABDOMINAL PAIN	NAUSEA AND VOMITING	DIARRHEA	GI BLEEDING	OBSTRUCTIVE JAUNDICE
Appendicitis	Medications	Infection	Ulcer disease	Bile duct stones
Gallstone disease	GI obstruction	Poorly absorbed sugars	Esophagitis	Cholangiocarcinoma
Pancreatitis	Motor disorders	Inflammatory bowel disease	Varices	Cholangitis
Diverticulitis	Functional bowel disorder	Microscopic colitis	Vascular lesions	Sclerosing cholangitis
Ulcer disease	Cyclic vomiting syndrome	Functional bowel disorder	Neoplasm	Ampullary stenosis
Esophagitis	Cannabinoid hyperemesis syndrome	Celiac disease	Diverticula	Ampullary carcinoma
GI obstruction	Enteric infection	Pancreatic insufficiency	Hemorrhoids	Pancreatitis
Inflammatory bowel disease	Pregnancy	Hyperthyroidism	Fissures	Pancreatic tumor
Functional bowel disorder	Endocrine disease	Ischemia	Inflammatory bowel disease	
Vascular disease	Motion sickness	Endocrine tumor	Infectious colitis	
Gynecologic causes	Central nervous system disease			
Renal stone				

mesenteric ischemia, and neoplasia. The most common causes of abdominal pain are IBS and functional dyspepsia.

Heartburn Heartburn, a burning substernal sensation, is reported intermittently by 40% of the population. Classically, heartburn results from excess gastroesophageal acid reflux, but some cases exhibit normal esophageal acid exposure and are caused by reflux of nonacidic material or heightened sensitivity of esophageal nerves.

Nausea and Vomiting Nausea and vomiting are caused by GI diseases, medications, toxins, infection, endocrine disorders, labyrinthine conditions, and central nervous system disease. Mechanical obstructions of the upper gut are commonly excluded as causes of chronic nausea and vomiting, but disorders of propulsion including gastroparesis and intestinal pseudoobstruction elicit similar symptoms. Nausea and vomiting also are commonly reported by patients with IBS and functional disorders of the upper gut (including chronic nausea vomiting syndrome, cyclic vomiting syndrome, and cannabinoid hyperemesis syndrome).

Altered Bowel Habits Altered bowel habits are common complaints in GI disease. Constipation may be reported as infrequent defecation, straining with defecation, passage of hard stools, or a sense of incomplete fecal evacuation and is caused by obstruction, motor disorders, medications, and endocrine diseases such as hypothyroidism and hyperparathyroidism. Diarrhea may be reported as frequent defecation, passage of loose or watery stools, fecal urgency, or a similar sense of incomplete evacuation. The differential diagnosis of diarrhea includes infections, inflammatory causes, malabsorption, and medications. IBS produces constipation, diarrhea, or an alternating bowel pattern. Fecal mucus is common in IBS, whereas pus and blood characterize IBD. Steatorrhea develops with malabsorption.

GI Bleeding Hemorrhage may develop from any gut organ. Upper GI bleeding presents with melena or hematemesis, whereas lower GI bleeding produces passage of bright red or maroon stools. However, briskly bleeding upper sites can elicit voluminous red rectal bleeding, whereas slowly bleeding ascending colon sites may produce melena. Chronic occult GI bleeding may present with iron deficiency anemia. Causes of upper GI bleeding include ulcer disease, gastroduodenitis, esophagitis, portal hypertensive etiologies, malignancy, tears across the gastroesophageal junction, and vascular lesions. Lower GI sources of hemorrhage include hemorrhoids, anal fissures, diverticula, ischemic colitis, neoplasm, IBD, infectious colitis, drug-induced colitis, arteriovenous malformations, and other vascular lesions.

Jaundice Jaundice results from prehepatic, intrahepatic, or posthepatic disease. Posthepatic causes of jaundice include biliary diseases, such as choledocholithiasis, acute cholangitis, primary sclerosing cholangitis, other strictures, and neoplasm, and pancreatic disorders, such as acute and chronic pancreatitis, stricture, and malignancy.

Other Symptoms Other symptoms are manifestations of GI disease. Dysphagia, odynophagia, and unexplained chest pain suggest esophageal disease. A globus sensation is reported with esophagopharyngeal conditions, but also occurs with functional GI disorders. Weight loss, anorexia, and fatigue present with neoplastic, inflammatory, motility, pancreatic, and psychiatric conditions. IBD is associated with hepatobiliary dysfunction, skin and eye lesions, and arthritis. Celiac disease may present with dermatitis herpetiformis. Jaundice can produce pruritus. Conversely, systemic diseases have GI consequences. Systemic lupus may cause gut ischemia, presenting with pain or bleeding. Severe burns may lead to gastric ulcer formation.

EVALUATION OF THE PATIENT WITH GI DISEASE

Evaluation of the patient with suspected GI disease begins with a careful history and examination. Subsequent investigation with tools to test gut structure or function and luminal constituents is indicated in selected cases. In patients with normal findings on diagnostic testing, validated symptom profiles are used to confidently diagnose a functional bowel disorder.

HISTORY

The history in suspected GI disease has several components. Symptom timing, patterns, and duration suggest specific etiologies. Short-duration symptoms commonly result from acute infection or inflammation, toxin exposure, or ischemia. Long-standing symptoms point to chronic inflammation, neoplasia, or functional bowel disorders. Luminal obstruction can present with dysphagia, nausea and vomiting, bloating and distention, or constipation depending on the site of blockage. Symptoms from mechanical obstruction, ischemia, IBD, and functional bowel disorders are worsened by meals, while ulcer symptoms may be relieved by eating or antacids. Ulcer pain occurs intermittently over weeks to months, whereas biliary colic has a sudden onset and lasts up to several hours. Acute pancreatitis pain is severe and persists for days to weeks. Meals elicit diarrhea while defecation relieves discomfort in some cases of IBD and IBS. Functional bowel disorders are exacerbated by stress. Sudden awakening from sound sleep by pain suggests organic rather than functional disease. Diarrhea from malabsorption usually improves with fasting, whereas secretory diarrhea persists without oral intake.

Symptom relation to other factors narrows the list of diagnostic possibilities. Obstructive symptoms with prior abdominal surgery raise concern for adhesions. Loose stools after gastrectomy or cholecystectomy suggest dumping syndrome or postcholecystectomy diarrhea. Symptom onset after travel prompts consideration of infection. Medications produce pain, altered bowel habits, or GI bleeding. Celiac disease is prevalent in people of northern European descent, whereas IBD is more common in Jewish populations. A sexual history may raise concern for infection or immunodeficiency.

Working groups have devised symptom criteria to improve diagnosis of functional bowel disorders to minimize the numbers of unnecessary diagnostic tests performed. The best accepted symptom-based criteria are the Rome criteria, which exhibit sensitivities and specificities of only 55–75% when tested against structural findings in IBS and functional dyspepsia, indicating a need for careful test selection in patients at high risk of organic disease.

PHYSICAL EXAMINATION

The physical examination complements information from the history. Abnormal vital signs provide diagnostic clues and determine the need for acute intervention. Fever suggests inflammation or neoplasm. Orthostasis is produced by significant blood loss, dehydration, sepsis, or autonomic neuropathy. Skin, eye, or joint findings may point to specific diagnoses. Neck examination with swallowing assessment evaluates dysphagia. Lung and cardiac examinations evaluate for cardiopulmonary disease as causes of abdominal pain or nausea. Pelvic examination tests for a gynecologic source of abdominal pain. Rectal examination may detect blood, indicating mucosal injury or neoplasm or a palpable inflammatory mass in appendicitis. Metabolic conditions and gut motor disorders have associated peripheral neuropathy.

Abdominal inspection may reveal distention from obstruction, tumor, or ascites or vascular abnormalities with liver disease. Ecchymoses develop with severe pancreatitis. Auscultation detects bruits or friction rubs from vascular disease or hepatic tumors. Loss of bowel sounds signifies ileus, whereas high-pitched, hyperactive sounds characterize intestinal obstruction. Percussion assesses liver size and detects shifting dullness from ascites. Palpation assesses for hepatosplenomegaly and neoplastic or inflammatory masses. Intestinal ischemia elicits severe pain but little tenderness. Patients with visceral pain may exhibit generalized discomfort, whereas those with parietal pain or peritonitis have localized pain with involuntary guarding, rigidity, or rebound. Patients with musculoskeletal abdominal wall pain may note tenderness exacerbated by Valsalva or leg lift maneuvers.

TOOLS FOR PATIENT EVALUATION

Laboratory, radiographic, and functional tests assist in diagnosis of suspected GI disease. The GI tract also is amenable to internal evaluation using endoscopy and to examination of luminal contents. Histopathologic examinations of GI tissues complement these tests.

Laboratory Laboratory tests facilitate diagnosis of GI disease. Iron-deficiency anemia suggests mucosal blood loss, whereas vitamin B₁₂ deficiency results from intestinal, gastric, or pancreatic disease. Either can result from inadequate oral intake. Leukocytosis and increased sedimentation rates and C-reactive proteins are found in inflammation, whereas leukopenia is seen in viremic illness. Severe vomiting or diarrhea elicits electrolyte disturbances, acid-base abnormalities, and elevated blood urea nitrogen. Pancreaticobiliary or liver disease produces elevated pancreatic or liver chemistries. Thyroid chemistries and cortisol and calcium levels evaluate for endocrinologic causes of symptoms. Pregnancy testing is considered for women with unexplained nausea. Serologic tests screen for celiac disease, IBD, connective tissue diseases, and paraneoplastic dysmotility syndromes. Hormone levels are obtained for suspected endocrine neoplasia. Intraabdominal malignancies produce tumor markers including the carcinoembryonic antigen CA 19-9 and α-fetoprotein. Blood testing also monitors medication therapy, as with thiopurine metabolite levels in IBD. Pharmacogenetic methods are being adopted to determine optimal patient populations for GI medication use. In conditions including IBD, research into novel biomarkers is being conducted to predict longitudinal course and treatment response. Other body fluids are sampled under certain circumstances. Ascitic fluid is analyzed for infection, malignancy, or findings of portal hypertension. Urine samples screen for carcinoid, porphyria, and heavy metal intoxication.

Luminal Contents Luminal contents can provide diagnostic clues. Stool samples are cultured for bacterial pathogens, examined for leukocytes and parasites, or tested for *Giardia* antigen. Duodenal aspirates can be examined for parasites or cultured for bacterial overgrowth. Fecal fat is quantified in possible malabsorption. Fecal elastase can be decreased with exocrine pancreatic insufficiency. Elevated fecal calprotectin or lactoferrin is found in inflammatory conditions such

as IBD. Stool electrolytes can be measured in diarrheal conditions. Laxative screens are performed for suspected laxative abuse. Fecal immunochemical and DNA tests have assumed roles in colon cancer screening in low-risk populations. Gastric acid is quantified to exclude gastrinoma. Esophageal pH/impedance testing is done for refractory symptoms of gastroesophageal reflux.

Endoscopy The gut is accessible to endoscopy, which can diagnose causes of bleeding, pain, nausea and vomiting, weight loss, altered bowel function, and fever. Table 321-2 lists common indications for endoscopic procedures. Upper endoscopy evaluates the esophagus, stomach, and duodenum, whereas colonoscopy assesses the colon and distal ileum. Upper endoscopy is advocated as the initial test for suspected ulcer disease, esophagitis, neoplasm, malabsorption, and Barrett's metaplasia because of its abilities to visualize and biopsy any abnormality. Colonoscopy is the preferred procedure for colon cancer screening and surveillance and to biopsy colitis or ileitis secondary to IBD, infection, ischemia, and radiation. Sigmoidoscopy examines the colon to the splenic flexure and excludes distal causes of bleeding, inflammation, or obstruction in young patients not at significant risk for colon cancer. For elusive GI bleeding from arteriovenous malformations or superficial ulcers, small-intestinal examination is performed with push enteroscopy, capsule endoscopy, or double-balloon enteroscopy. Capsule endoscopy also visualizes small-intestinal Crohn's disease in individuals with negative radiography. Endoscopic ultrasound (EUS) diagnoses and stages GI malignancy, excludes choledocholithiasis, evaluates pancreatitis, and assesses anal continuity. Endoscopic retrograde cholangiopancreatography (ERCP) provides diagnoses of pancreatic and biliary disease.

The development of novel imaging protocols permits optical biopsies to define mucosal histology and detect dysplasia in selected settings. Methods employed include narrow-band imaging,

TABLE 321-2 Common Indications for Endoscopy

UPPER ENDOSCOPY	COLONOSCOPY	ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY	ENDOSCOPIC ULTRASOUND	CAPSULE ENDOSCOPY	DOUBLE-BALLOON ENDOSCOPY
Dyspepsia despite treatment	Cancer screening	Jaundice	Staging of malignancy	Obscure bleeding	Ablation of small-intestinal bleeding sources
Dyspepsia with signs of organic disease	Lower gastrointestinal (GI) bleeding	Postbiliary surgery complaints	Characterize and biopsy submucosal mass	Suspected Crohn's disease of the small intestine	Biopsy of suspicious small-intestinal masses/ulcers
Refractory vomiting	Anemia	Cholangitis	Bile duct stones		
Dysphagia	Diarrhea	Gallstone pancreatitis	Chronic pancreatitis		
Upper GI bleeding	Polypectomy	Pancreatic/biliary/ampullary tumor	Drain pseudocyst		
Anemia	Obstruction	Unexplained pancreatitis	Anal continuity		
Weight loss	Biopsy radiologic abnormality	Pancreatitis with unrelenting pain	Direct stent placement		
Malabsorption	Cancer surveillance: family history prior polyp/cancer, colitis	Fistulas			
Biopsy radiologic abnormality		Biopsy radiologic abnormality			
Polypectomy	Palliate neoplasm	Pancreaticobiliary drainage			
Place gastrostomy	Remove foreign body	Sample bile			
Barrett's surveillance	Place stent across stenosis	Sphincter of Oddi manometry			
Palliate neoplasm					
Sample duodenal tissue/fluid					
Remove foreign body					
Endoscopic mucosal resection or endoscopic submucosal dissection for dysplasia or early cancer					
Place stent across stenosis					
Endoscopic myotomy for achalasia or gastroparesis					
Endoscopic bariatric procedures					

chromoendoscopy, confocal laser endomicroscopy, and optical coherence tomography in colitis, Barrett's esophagus, and gastric cancer surveillance. Artificial intelligence using machine learning techniques shows promise in detecting dysplasia and early cancer in still images from biopsy tissues.

Radiography/Nuclear Medicine Radiographic tests evaluate gut diseases and extraluminal structures. Contrast radiography with barium provides mucosal definition and can assess gut transit and pelvic floor dysfunction. An esophagram is the initial procedure to exclude subtle rings, strictures, or achalasia as causes of dysphagia, whereas small-bowel contrast radiology detects intestinal tumors, strictures, and fistulae and can estimate intestinal transit. Contrast enemas are performed when colonoscopy is unsuccessful or contraindicated. Ultrasound and computed tomography (CT) evaluate regions not accessible by endoscopy or contrast studies, including the liver, pancreas, gallbladder, kidneys, and retroperitoneum, and are useful for diagnosing mass lesions, fluid collections, organ enlargement, and, in the case of ultrasound, gallstones. CT and magnetic resonance (MR) colonography have been considered as alternatives to colonoscopy for colon cancer screening but have not commonly been adopted. MR methods image the pancreaticobiliary ducts to exclude neoplasm, stones, and sclerosing cholangitis and the liver to characterize benign and malignant tumors. Specialized CT or MR enterography quantifies IBD intensity. Angiography excludes mesenteric ischemia and determines spread of malignancy. Angiographic techniques also access the biliary tree in obstructive jaundice. CT and MR techniques screen for mesenteric occlusion, thereby limiting exposure to angiographic dyes. Positron emission tomography (PET) can distinguish malignant from benign disease in several organ systems. Imaging with DOTA-octreotate and related agents has improved detection of neuroendocrine tumors by combined PET-CT techniques.

Scintigraphy evaluates structural abnormalities and quantifies luminal transit. Radionuclide scans localize bleeding sites in patients with brisk hemorrhage to direct therapy with endoscopy, angiography, or surgery. Radiolabeled leukocyte scans search for intraabdominal abscesses not visualized on CT. Biliary scintigraphy complements ultrasound in assessing for cholecystitis. Scintigraphy to quantify esophageal and gastric emptying is well established, whereas techniques to measure small-intestinal or colonic transit are less widely used.

Histopathology Endoscopic mucosal biopsies evaluate for inflammatory, infectious, and neoplastic disease. Deep rectal biopsies facilitate diagnosis of Hirschsprung's disease or amyloid. Liver biopsy is performed for abnormal liver chemistries, unexplained jaundice, and some cases of viral hepatitis, and following liver transplant to exclude rejection. Biopsies obtained during CT or ultrasound evaluate for intraabdominal conditions not accessible by endoscopy.

Functional Testing Tests of gut function provide important data when structural testing is nondiagnostic. Functional testing of motor activity is provided by newer high-resolution manometric techniques. Esophageal manometry is useful for suspected achalasia, whereas small-intestinal manometry tests for pseudoobstruction and colon manometry evaluates for colonic inertia. A wireless motility capsule measures transit and contractile activity in the stomach, small intestine, and colon in a single test. Anorectal manometry with balloon expulsion testing is used for unexplained incontinence or constipation from outlet dysfunction. Biliary manometry tests for sphincter of Oddi dysfunction with unexplained biliary pain. The endoluminal functional lumen imaging probe can measure reduced distensibility in the lower esophageal sphincter in achalasia, the pylorus in gastroparesis, and the anus for defecation disorders. Measurement of breath hydrogen while fasting and after oral mono- or oligosaccharide challenge can screen for carbohydrate intolerance and small-intestinal bacterial overgrowth. Urea breath testing assesses for persistent *Helicobacter pylori* infection, while gastric emptying breath testing is an alternative to scintigraphy for gastroparesis diagnosis.

TREATMENT

Gastrointestinal Disease

Management options for GI diseases depend on the cause of symptoms. Available treatments include modifications of dietary intake, medications, treatment of gut dysbiosis, luminal intubation, interventional endoscopy or radiology techniques, surgery, psychological approaches, and physical therapy. Given the hereditary predisposition of many GI diseases, genetic testing may be indicated in some patients. Improved smartphone applications are being adopted for diverse purposes ranging from providing instructions for endoscopy preparation to educating and promoting adherence to diet restrictions in several disorders.

NUTRITIONAL MANIPULATION

Dietary modifications for GI disease include those that only reduce symptoms, those that correct pathologic defects, or those that replace normal food intake with enteral or parenteral formulations. Changes that improve symptoms but do not reverse organic abnormalities include lactose restriction for lactase deficiency, liquid meals in gastroparesis, carbohydrate restrictions with dumping syndrome, and low-FODMAP (fermentable oligo-di-monosaccharides and polyols) diets in IBS. The gluten-free diet for celiac disease exemplifies a primary therapy to reduce mucosal inflammation. Likewise, elimination diets may improve histology and symptoms in some cases of eosinophilic esophagitis. Medium-chain triglycerides replace normal fats in short-gut syndrome or severe ileal disease. Perfusion liquid meals through a gastrostomy is performed in those who cannot swallow safely. Enteral jejunostomy feedings are considered for gastric dysmotility syndromes that preclude feeding into the stomach. Intravenous hyperalimentation is used for generalized gut malfunction, which does not permit enteral nutrition.

PHARMACOTHERAPY

Several medications can treat GI diseases. Considerable resources are expended on over-the-counter remedies. Many prescription drug classes are offered as short-term or continuous therapy of GI illness. Alternative treatments are popular in conditions for which traditional therapies provide incomplete relief.

Over-the-Counter Agents Over-the-counter agents are reserved for mild GI symptoms. Antacids, histamine H₂ antagonists, and proton pump inhibitors (PPIs) decrease symptoms in gastroesophageal reflux disease (GERD) and dyspepsia. Fiber supplements, stool softeners, enemas, and laxatives are used for constipation. Laxatives are categorized as stimulants, osmotic agents (including isotonic preparations containing polyethylene glycol), and poorly absorbed sugars. Nonprescription antidiarrheal agents include bismuth subsalicylate, kaolin-pectin combinations, and loperamide, whereas lactase enzyme pills are used for lactose intolerance. Gaseous symptoms may be reduced by bacterial α -galactosidase, antiflatulents, and adsorbents. In general, using a nonprescription preparation for more than a short time for chronic persistent symptoms should be supervised by a health care provider.

Prescription Drugs Prescription drugs are approved for a broad range of GI diseases. Higher-dose prescription PPIs are advocated for GERD when over-the-counter preparations are inadequate. Cytoprotective agents are available for upper gut ulcers but are less frequently prescribed. Prokinetic drugs stimulate GI propulsion in gastroparesis, pseudoobstruction, and slow-transit constipation. Secretagogue drugs are prescribed for constipation refractory to other agents, whereas peripheral opioid antagonists are offered for opioid-induced constipation. Prescription antidiarrheals include opioid drugs, anticholinergic antispasmodics, tricyclics, bile acid binders, and serotonin antagonists. Antispasmodics and anti-depressants also are useful for functional GI disorders, whereas narcotics are used for pain control in organic conditions such as disseminated malignancy and chronic pancreatitis. Antiemetics

reduce nausea and vomiting. Potent pancreatic enzymes decrease malabsorption and pain from pancreatic disease. Antisecretory drugs such as the somatostatin analogue octreotide treat hypersecretory states. Some functional GI disorders require use of neuro-modulators, including tricyclic agents, for control of pain, diarrhea, or nausea. Antibiotics treat *H. pylori*-induced ulcers, infectious diarrhea, diverticulitis, intestinal bacterial overgrowth, and Crohn's disease. Anti-inflammatory and immunomodulatory drugs are used in IBD, microscopic colitis, refractory celiac disease, autoimmune pancreatitis, and gut vasculitis. Over the past decade, several newer biologic agents, including agents that inhibit tumor necrosis activity, other proinflammatory cytokines, and Janus kinase signaling or serve as antiadhesion molecules, have had dramatic impact in Crohn's disease and ulcerative colitis. Biologics that deplete eosinophils or inhibit mast cells show promise in eosinophilic disorders of the gut. Chemotherapy with or without radiotherapy is offered for GI malignancies. Most GI carcinomas respond poorly to such therapy, whereas lymphomas may be cured with such intervention.

Complementary and Alternative Medicine Treatments Alternative treatments are marketed to treat GI symptoms. Ginger, acupuncture, and acustimulation have been advocated for nausea, whereas pyridoxine has been investigated for nausea of first-trimester pregnancy. Peppermint oil and caraway seed oil products and herbal preparations such as STW 5 (a mixture of nine herbs) are useful in cases of functional dyspepsia and IBS. Low-potency pancreatic enzyme preparations are sold as general digestive aids but have little evidence to support their efficacy.

THERAPIES TARGETING GUT DYSBIOSIS

Some cases of diarrhea-predominant IBS respond to nonabsorbable antibiotics. Oral antibiotics also are the mainstay of managing intestinal bacterial overgrowth. Probiotics containing active bacterial cultures and prebiotics that selectively nourish nonnoxious commensal bacteria are used as adjuncts in some cases of infectious diarrhea and IBS, with limited evidence of efficacy. Transplantation of donor feces into the colon by colonoscopy or enema is effective treatment for recurrent, refractory *Clostridium difficile* colitis, and numerous trials are being conducted to assess utility of this technique in IBS, IBD, and liver disease.

LUMINAL SUCTION AND LAVAGE

Nasogastric tube suction decompresses the upper gut in ileus or mechanical obstruction. Nasogastric lavage of saline or water in the patient with upper GI hemorrhage determines the rate of bleeding and helps evacuate blood before therapeutic endoscopy. Enteral feedings can be delivered through nasogastric or nasoenteric tubes. Enemas relieve fecal impaction or assist in gas evacuation in acute colonic pseudoobstruction. A rectal tube can be placed to vent the distal colon in colonic pseudoobstruction and other colonic distension disorders.

INTERVENTIONAL ENDOSCOPY AND RADIOLGY

In addition to its diagnostic role, endoscopy has numerous therapeutic capabilities. Cautery techniques and injection of vasoconstrictor substances can stop hemorrhage from ulcers and vascular malformations. Endoscopic encirclement of varices and hemorrhoids with constricting bands stops hemorrhage from these sites, whereas endoscopically placed clips can occlude arterial bleeding sites. Endoscopically delivered hemostatic powder sprays are approved to stop brisk GI bleeding. Endoscopy can remove polyps or debulk lumen-narrowing malignancies. Colonoscopy can withdraw excess gas in some cases of acute colonic pseudoobstruction. Endoscopic mucosal resection, submucosal dissection, and radiofrequency techniques can ablate some cases of Barrett's esophagus with dysplasia or superficial cancer and early gastric malignancies. Obstructions of the gut lumen and pancreaticobiliary tree are relieved by endoscopic dilation or placing plastic or expandable metal stents. Endoscopic sphincterotomy of the ampulla of Vater relieves symptoms of choledocholithiasis. Cholangioscopy can help

with stone lithotripsy in the common bile duct, ablation of small ductal tumors, and placement of gallbladder stents to facilitate drainage in nonoperative candidates. Methods employing interventional EUS have been developed for pancreatic cyst gastrostomy using lumen-apposing metal stents, pancreatic necrosectomy, and placement of fiducial markers to direct pancreatic and rectal radiotherapy. EUS also has been used to facilitate endoscopic access to the excluded distal stomach in patients who have undergone bariatric gastric bypass surgery using similar stents so that ERCP can be done for pancreaticobiliary conditions. Likewise, EUS-directed stent placement can manage postsurgical stenoses after pancreatic resection. Endoscopy is sometimes used to insert gastric feeding tubes. Peroral endoscopic myotomy is therapeutically performed on the lower esophageal sphincter in achalasia and on the pylorus in gastroparesis. Endoscopic treatments for acid reflux including radiofrequency therapy, transoral fundoplication, endoscopic stapling, and antireflux mucosectomy have been developed, but many offer unproven utility. Endoscopic bariatric methodologies, including intragastric balloons, endoscopic sleeve gastoplasty, and duodenal resurfacing and diversion, have been introduced.

Radiologic measures also are useful in GI disease. Angiographic embolization or vasoconstriction decreases bleeding from gut sites not amenable to endoscopic intervention. Dilatation or stenting with fluoroscopic guidance relieves luminal strictures. Contrast enemas can reduce volvulus and evacuate air in acute colonic pseudoobstruction. CT and ultrasound help drain abdominal fluid collections, in many cases obviating the need for surgery. Percutaneous transhepatic cholangiography relieves biliary obstruction when ERCP is contraindicated. Transjugular intrahepatic portosystemic shunts are commonly performed by interventional radiologists for variceal hemorrhage not amenable to endoscopic therapy. Lithotripsy can fragment gallstones in patients who are not candidates for surgery. Radiologic approaches are often chosen over endoscopy for gastroenterostomy placement. Finally, central venous catheters for parenteral nutrition may be placed using radiographic techniques.

SURGERY

Surgery is performed to cure disease, control symptoms, maintain nutrition, or palliate unresectable neoplasm. Surgery can cure medication-unresponsive ulcerative colitis, diverticulitis, cholecystitis, appendicitis, and intraabdominal abscess, but can only reduce symptoms and treat disease complications in Crohn's disease. Surgery is mandated for ulcer complications such as bleeding, obstruction, or perforation and intestinal obstructions that persist after conservative care. Gastroesophageal fundoplication is performed for severe ulcerative esophagitis and drug-refractory symptomatic acid reflux. Achalasia responds to operations to reduce lower esophageal sphincter tone. Operations for motor disorders include implanted electrical stimulators for gastroparesis and electrical devices and artificial sphincters for fecal incontinence. Surgery may be needed to place a jejunostomy for long-term enteral feedings.

PSYCHOLOGICAL APPROACHES AND PHYSICAL THERAPY

Psychological therapies, including psychotherapy, cognitive behavioral therapy, and hypnosis, have shown efficacy in functional bowel disorders. Patients with significant psychological dysfunction and those with little response to treatments targeting the gut are likely to benefit from this form of therapy. Biofeedback methods administered by physical therapies are accepted for treating refractory fecal incontinence or constipation secondary to dyssynergia.

FURTHER READING

- A HR et al: AGA Clinical Practice Update on pancreas cancer screening in high-risk individuals: An expert review. *Gastroenterology* 159:358, 2020.
- B JS et al: Major trends in gastroenterology and hepatology between 2010 and 2019: An overview of advances from the past

decade selected by the editorial board of the American Journal of Gastroenterology. Am J Gastroenterol 115:1007, 2020.

G S et al: Recommendations for follow-up after colonoscopy and polypectomy: A consensus update by the US Multi-Society Task Force on Colorectal Cancer. Gastroenterology 158:1131, 2020.

K L et al: Best practice update: Incorporating psychogastroenterology into management of digestive disorders. Gastroenterology 154:1249, 2018.

L C A et al: British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. Gut 68:s1, 2019.

O V et al: The gut-brain axis and the microbiome: Mechanisms and clinical implications. Clin Gastroenterol Hepatol 17:322, 2019.

322

Gastrointestinal Endoscopy

Louis Michel Wong Kee Song, Mark Topazian

Gastrointestinal endoscopy has been attempted for >200 years, but the introduction of semirigid and flexible gastrosopes in the mid-twentieth century marked the dawn of the modern endoscopic era. Since then, rapid advances in endoscopic technology have led to dramatic changes in the diagnosis and treatment of many digestive diseases. Innovative endoscopic devices and new endoscopic treatment modalities continue to expand the use of endoscopy in patient care.

Flexible endoscopes provide an electronic video image generated by a charge-coupled device (CCD) or a complementary metal oxide semiconductor (CMOS) chip in the tip of the endoscope. Operator controls permit deflection of the endoscope tip; fiberoptic bundles or light-emitting diodes provide light at the tip of the endoscope; and working channels allow washing, suctioning, and the passage of instruments (Fig. 322-1). Progressive changes in the diameter and stiffness of endoscopes have improved the ease and patient tolerance of endoscopy. High-resolution and high-definition endoscopes equipped with electronic and optical magnification capabilities enable acquisition of images with a high level of detail. Advanced imaging techniques, including narrow-band imaging (Fig. 322-2) and real-time



FIGURE 322-1 Gastrointestinal endoscope. Shown here is a conventional colonoscope with control knobs for tip deflection, push buttons for suction and air insufflation (single arrows), and a working channel for passage of accessories (double arrows).



A



B

FIGURE 322-2 Flat colon polyp. A. White-light imaging. B. Corresponding narrow-band imaging enhances mucosal features and lesion delineation.

image-processing enhancement algorithms, aid in tissue characterization or differentiation.

ENDOSCOPIC PROCEDURES

■ UPPER ENDOSCOPY

Upper endoscopy, also referred to as esophagogastroduodenoscopy (EGD), is performed by passing a flexible endoscope through the mouth into the esophagus, stomach, and duodenum. The procedure is the best method for examining the upper gastrointestinal mucosa (Fig. 322-3). While the upper gastrointestinal radiographic series has similar accuracy for diagnosis of duodenal ulcer (Fig. 322-4), EGD is superior for detection of gastric ulcers (Fig. 322-5) and flat mucosal lesions, such as Barrett's esophagus (Fig. 322-6), and it permits directed biopsy and endoscopic therapy. Intravenous sedation is given to most patients in the United States to ease the anxiety and discomfort of the procedure, although in many countries, EGD is routinely performed with topical pharyngeal anesthesia only. Patient tolerance of unsedated EGD is improved by the use of an ultrathin, 5-mm diameter endoscope that can be passed transorally or transnasally.

■ COLONOSCOPY

Colonoscopy is performed by passing a flexible colonoscope through the anal canal into the rectum and colon. The cecum is reached in



A



B



C



D



E

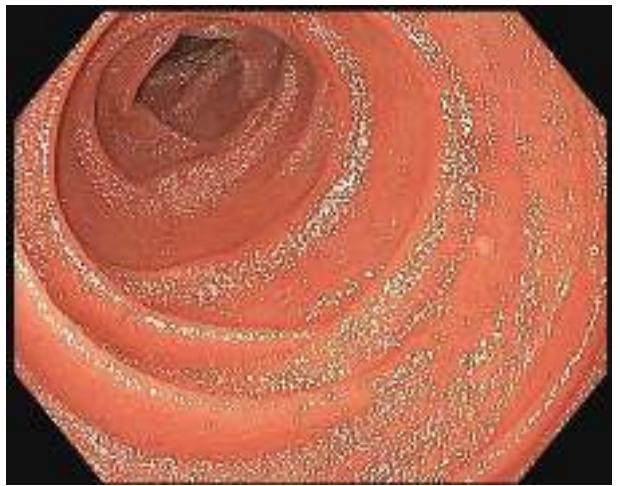


F

FIGURE 322-3 Normal upper endoscopic examination. A. Esophagus. B. Gastroesophageal junction. C. Gastric fundus. D. Gastric body. E. Gastric antrum. F. Pylorus. G. Duodenal bulb. H. Second portion of the duodenum.



G

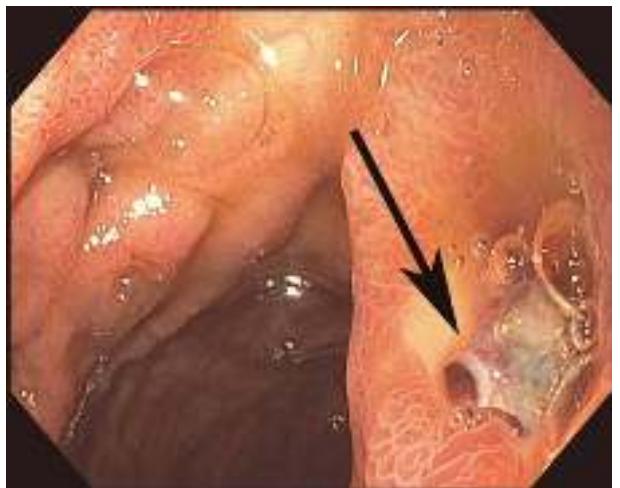


H

FIGURE 322-3 (Continued)



A



B

FIGURE 322-4 Duodenal ulcers. A. Ulcer with a small, flat, pigmented spot in its base. B. Ulcer with a visible vessel (arrow) in a patient with recent hemorrhage.



A



B

FIGURE 322-5 Gastric ulcers. A. Benign gastric ulcer in the antrum. B. Malignant gastric ulcer involving greater curvature of stomach.



FIGURE 322-6 Barrett's esophagus. *A*, Salmon-colored Barrett's mucosa extending proximally from the gastroesophageal junction. *B*, Barrett's esophagus with a suspicious nodule (arrow) identified during endoscopic surveillance. *C*, Histologic finding of intramucosal adenocarcinoma in the endoscopically resected nodule. Tumor extends into the esophageal submucosa (arrow). *D*, Barrett's esophagus with locally advanced adenocarcinoma.

>95% of cases, and the terminal ileum (Fig. 322-7) can often be examined. Colonoscopy is the gold standard for imaging the colonic mucosa (Fig. 322-8). Colonoscopy has greater sensitivity than barium enema for colitis (Fig. 322-9), polyps (Fig. 322-10), and cancer (Fig. 322-11). CT colonography rivals the accuracy of colonoscopy for detection of some polyps and cancer, although it is not as sensitive for the detection of flat lesions, such as serrated polyps (Fig. 322-12). Moderate sedation is usually given before colonoscopy in the United States, although a willing patient and a skilled examiner can complete the procedure without sedation in many cases.

■ FLEXIBLE SIGMOIDOSCOPY

Flexible sigmoidoscopy is akin to colonoscopy, but it visualizes only the rectum and a variable portion of the left colon, typically to 60 cm

from the anal verge. This procedure causes abdominal cramping, but it is brief and usually performed without sedation. Flexible sigmoidoscopy is primarily used for evaluation of diarrhea and rectal outlet bleeding.

■ SMALL BOWEL ENDOSCOPY

Three endoscopic techniques are currently used to evaluate the small intestine, most often in patients presenting with presumed small-bowel bleeding. For *capsule endoscopy*, the patient swallows a disposable capsule that contains a CMOS chip camera. Color still images (Fig. 322-13) are transmitted wirelessly to an external receiver at several frames per second until the capsule's battery is exhausted or it is passed into the toilet. Capsule endoscopy enables visualization of the small-bowel mucosa beyond the reach of a conventional

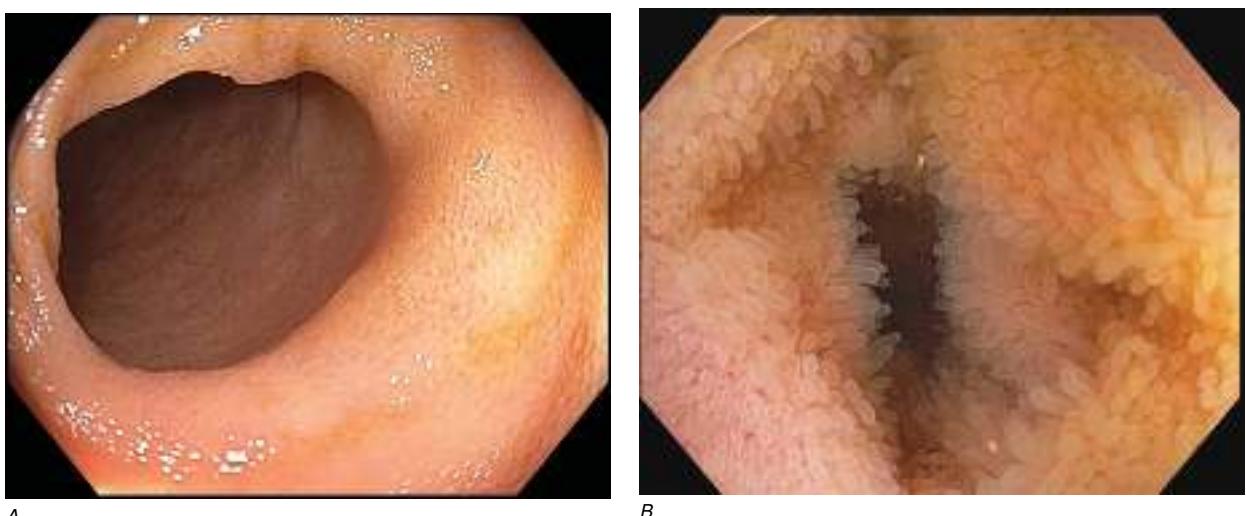


FIGURE 322-7 Colonoscopic view of terminal ileum. *A*, Normal-appearing terminal ileum (TI). *B*, View of normal villi of TI enhanced by examination under water immersion.

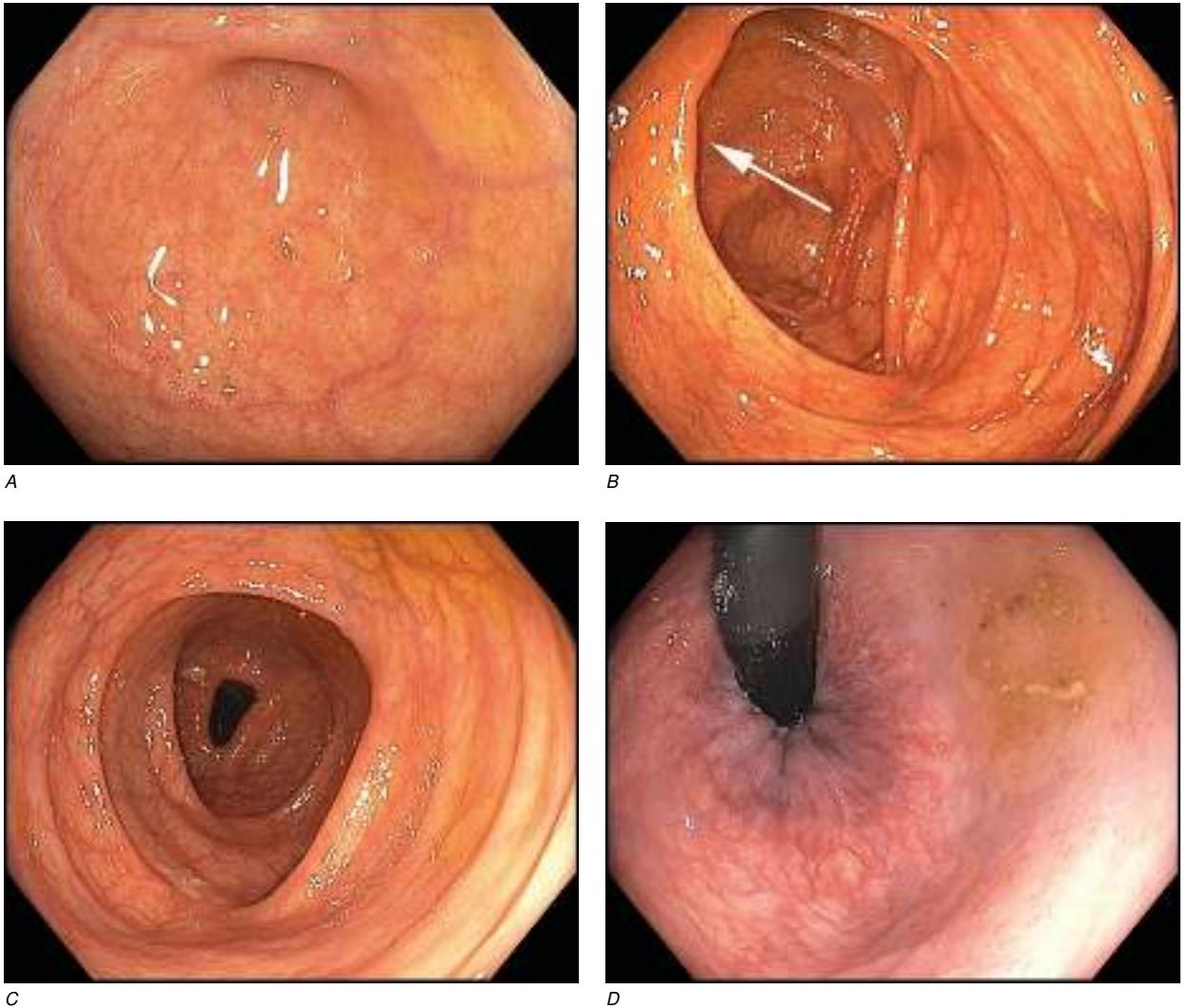


FIGURE 322-8 Normal colonoscopic examination. *A*. Cecum with view of appendiceal orifice. *B*. Ileocecal valve. *C*. Normal-appearing colon. *D*. Rectum (retroflexed view).

endoscope, and at present, it is solely a diagnostic procedure. Patients with a history of prior intestinal surgery or Crohn's disease are at risk for capsule retention at the site of a clinically unsuspected small-bowel stricture, and ingestion of a "patency capsule" composed of radiologically opaque biodegradable material may be indicated prior to capsule endoscopy in such patients.

Push enteroscopy is generally performed using a variable-stiffness pediatric or adult colonoscope or a dedicated enteroscope with or without the assistance of a stiffening overtube that extends from the mouth to the small intestine. The proximal to mid-jejunum is usually reached, and the instrument channel of the endoscope allows for biopsy or endoscopic therapy.

Deeper insertion into the small bowel can be accomplished by *device-assisted enteroscopy*, which may utilize inflatable balloons at the tip of the enteroscope and/or an overtube (*single- or double-balloon enteroscopy*) or a rotating, screw-like overtube (*motorized spiral enteroscopy*) to pleat the small intestine onto the endoscope (Fig. 322-14, Video V5-1). With device-assisted enteroscopy, the entire small intestine can be visualized in some patients when both the oral and anal routes of insertion are used. Biopsies and endoscopic therapy can be performed throughout the visualized small bowel (Fig. 322-15).

■ ENDOSCOPIC RETROGRADE CHOLANGIOPANCREATOGRAPHY

During endoscopic retrograde cholangiopancreatography (ERCP), a side-viewing endoscope is passed through the mouth to the duodenum, the ampulla of Vater is identified and cannulated with a thin plastic catheter, and radiographic contrast material is injected into the bile duct and pancreatic duct under fluoroscopic guidance (Fig. 322-16). When indicated, the major papilla can be incised using the technique of endoscopic sphincterotomy (Fig. 322-17). Stones can be retrieved from the ducts, biopsies can be performed, strictures can be dilated and/or stented (Fig. 322-18), and ductal leaks can be treated (Fig. 322-19). ERCP is usually performed for therapy but is also important diagnostically as it facilitates tissue sampling of biliary or pancreatic ductal strictures.

■ ENDOSCOPIC ULTRASOUND

Endoscopic ultrasound (EUS) utilizes ultrasound transducers incorporated into the tip of a flexible endoscope. Ultrasound images are obtained of the gut wall and adjacent organs, vessels, lymph nodes, and other structures. High-resolution images are obtained by bringing a high-frequency ultrasound transducer close to the area of interest via endoscopy. EUS provides the most accurate preoperative local staging



A



1



6



1

FIGURE 322-9 Causes of colitis. *A*. Chronic ulcerative colitis with diffuse ulcerations and exudates. *B*. Severe Crohn's colitis with deep ulcers. *C*. Pseudomembranous colitis with yellow, adherent pseudomembranes. *D*. Ischemic colitis with patchy mucosal edema, subepithelial hemorrhage, superficial ulcerations, and cyanosis.



A



1

FIGURE 322-10 Colonic polyps. A. Pedunculated polyp on a stalk. B. Sessile polyp



FIGURE 322-11 Ulcerated colon adenocarcinoma narrowing the colonic lumen.



A



B



C

FIGURE 322-12 Flat serrated polyp in the cecum. A. Appearance of the lesion under conventional white-light imaging. B. Mucosal patterns and boundary of the lesion enhanced with narrow-band imaging. C. Submucosal lifting of the lesion with dye (methylene blue) injection prior to resection.



FIGURE 322-13 Capsule endoscopy. Image of a jejunal vascular ectasia.

of esophageal, pancreatic, and rectal malignancies (Fig. 322-20), but it does not detect most distant metastases. EUS is also useful for diagnosis of bile duct stones, gallbladder disease, subepithelial gastrointestinal lesions, and chronic pancreatitis. Fine-needle aspirates and core biopsies of organs, masses, and lymph nodes in the posterior mediastinum, abdomen, retroperitoneum, and pelvis can be obtained under EUS guidance (Fig. 322-21). EUS-guided therapeutic procedures are increasingly performed, including drainage of abscesses, pseudocysts, and pancreatic necrosis into the gut lumen (Video V5-2); celiac plexus neurolysis for treatment of pancreatic pain; ethanol ablation of pancreatic neuroendocrine tumors; treatment of gastrointestinal hemorrhage; and drainage of obstructed biliary and pancreatic ducts.

■ NATURAL ORIFICE TRANSLUMINAL ENDOSCOPIC SURGERY

Natural orifice transluminal endoscopic surgery (NOTES) is an evolving collection of endoscopic methods that entail passage of an

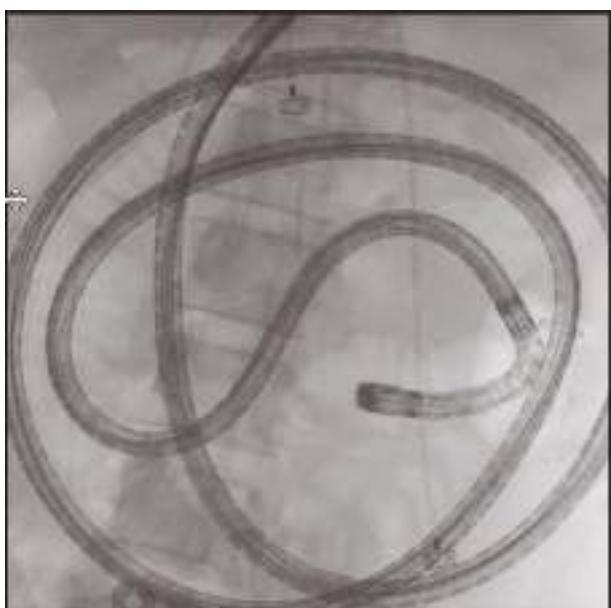
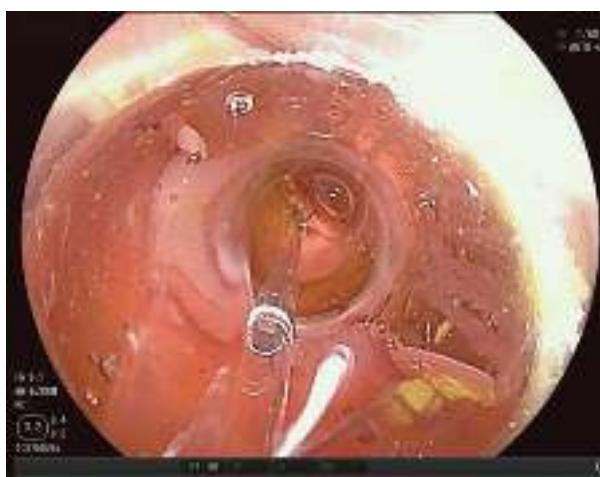


FIGURE 322-14 Double-balloon enteroscopy. Radiograph of the orally inserted instrument deep in the small intestine.



A



B



C

FIGURE 322-15 Nonsteroidal anti-inflammatory drug (NSAID)-induced proximal ileal stricture managed via double-balloon enteroscopy. A. High-grade ileal stricture causing obstructive symptoms. B. Balloon dilation of the ileal stricture. C. Appearance of the stricture after dilation.



A



B

FIGURE 322-16 Endoscopic retrograde cholangiopancreatography (ERCP) for bile duct stones with cholangitis. A. Faceted bile duct stones are demonstrated in the common bile duct. B. After endoscopic sphincterotomy, the stones are extracted with a Dormia basket. A small abscess communicates with the left hepatic duct.

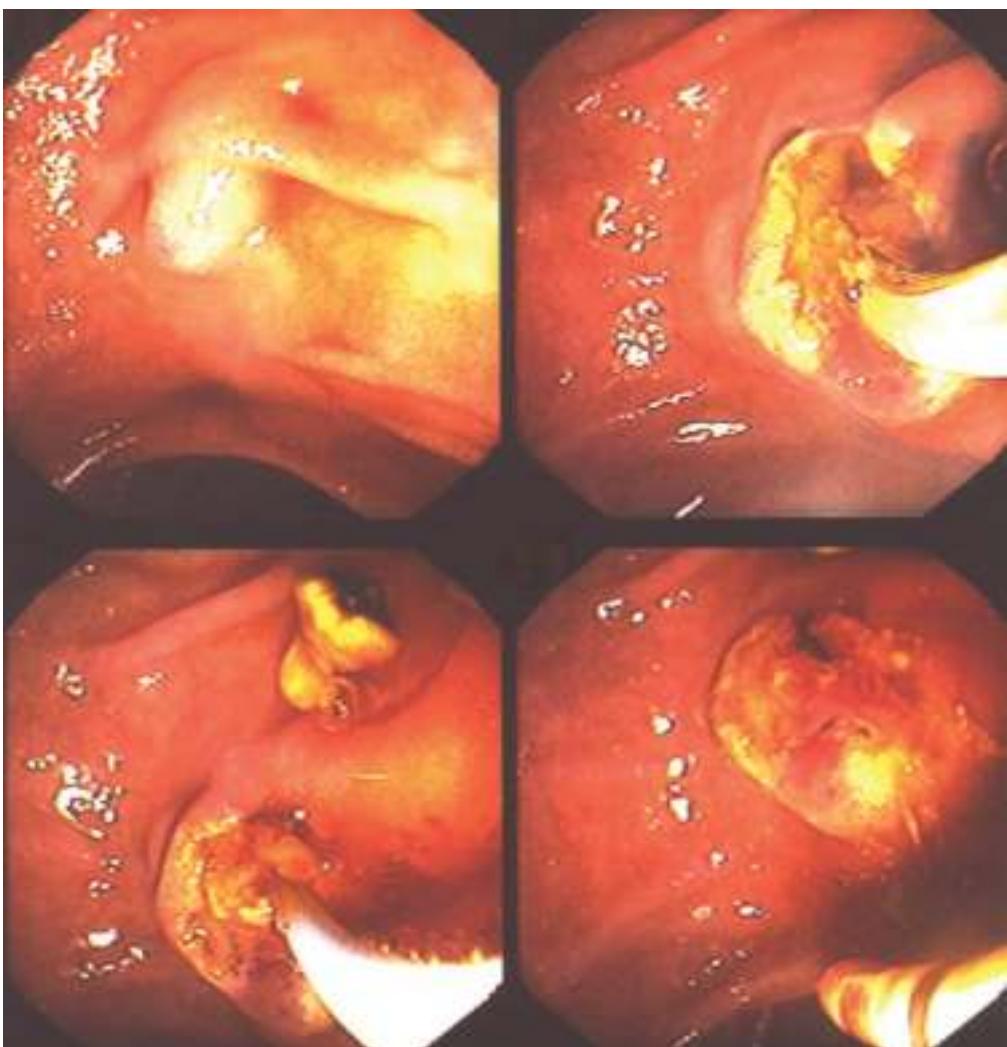


FIGURE 322-17 Endoscopic sphincterotomy. *A*, A normal-appearing ampulla of Vater. *B*, Sphincterotomy is performed with electrosurgery. *C*, Bile duct stones are extracted with a balloon catheter. *D*, Final appearance of the sphincterotomy.

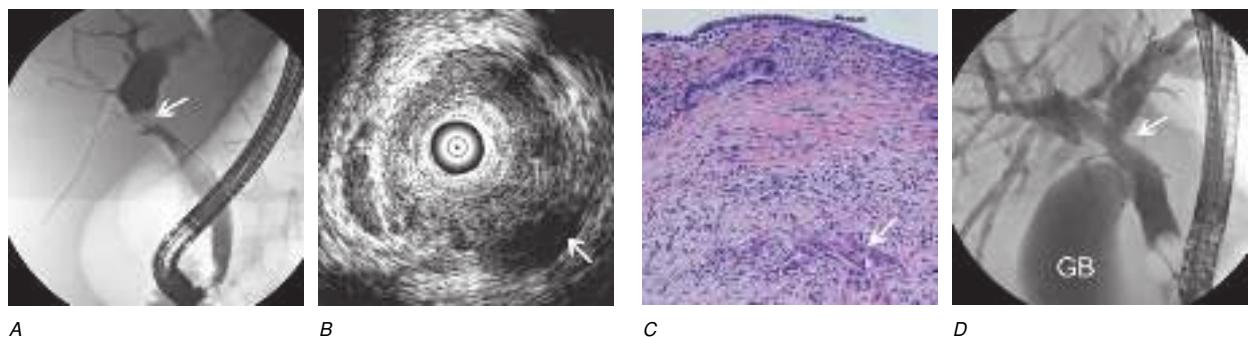


FIGURE 322-18 Endoscopic diagnosis, staging, and palliation of hilar cholangiocarcinoma. *A*, Endoscopic retrograde cholangiopancreatography (ERCP) in a patient with obstructive jaundice demonstrates a malignant-appearing stricture of the biliary confluence extending into the left and right intrahepatic ducts. *B*, Intraductal ultrasound of the biliary stricture demonstrates marked bile duct wall thickening due to tumor (*T*) with partial encasement of the hepatic artery (arrow). *C*, Intraductal biopsy obtained during ERCP demonstrates malignant cells infiltrating the submucosa of the bile duct wall (arrow). *D*, Endoscopic placement of bilateral self-expanding metal stents (arrow) relieves the biliary obstruction. *GB*, gallbladder. (Image courtesy of Dr. Thomas Smyrk.)



FIGURE 322-19 Bile leak (arrow) from a duct of Luschka after laparoscopic cholecystectomy. Contrast leaks from a small right intrahepatic duct into the gallbladder fossa, then flows into the pigtail of a percutaneous drainage catheter.

endoscope or its accessories into or through the wall of the gastrointestinal tract to perform diagnostic or therapeutic interventions. Some NOTES procedures, such as percutaneous endoscopic gastrostomy (PEG) or endoscopic necrosectomy of pancreatic necrosis, are well-established clinical procedures (Video V5-2); others such as peroral endoscopic myotomy (POEM) for achalasia (Fig. 322-22) and gastroparesis, peroral endoscopic tumorectomy (POET) (Fig. 322-23), and endoscopic full-thickness resection (EFTR) of gastrointestinal mural lesions (Fig. 322-24, Video V5-3), are emerging as minimally invasive therapeutic options. NOTES is an area of continuing innovation and endoscopic research.

■ ENDOSCOPIC RESECTION AND CLOSURE TECHNIQUES

Endoscopic mucosal resection (EMR) (Fig. 322-25, Video V5-4) and endoscopic submucosal dissection (ESD) (Fig. 322-26, Video V5-5) are the two commonly used techniques for the resection of benign and early-stage malignant gastrointestinal neoplasms. In addition to providing larger specimens for more accurate histopathologic assessment and diagnosis, these techniques can be potentially curative for some dysplastic lesions and focal intramucosal carcinomas involving

the esophagus, stomach, and colon. Several devices are available for closure of EMR and ESD defects, as well as gastrointestinal fistulas and perforations. Endoscopic clips deployed through the working channel of an endoscope have been used for many years to treat bleeding lesions, and the development of larger over-the-scope clips has facilitated endoscopic closure of gastrointestinal fistulas and perforations not previously amenable to endoscopic therapy (Video V5-6). Endoscopic suturing can be used to close some perforations and large defects (Fig. 322-27), anastomotic leaks, and fistulas. Other potential indications for endoscopic suturing include stent fixation to prevent migration (Fig. 322-28, Video V5-7) and endoscopic bariatric procedures. These technologies are playing an expanding role in patient care.

RISKS OF ENDOSCOPY

Medications used during moderate sedation may cause respiratory depression or allergic reactions. All endoscopic procedures carry some risk of bleeding and gastrointestinal perforation. The risk is small with diagnostic upper endoscopy, flexible sigmoidoscopy, and colonoscopy (<1:1000 procedures), but it ranges from 0.5 to 5% when therapeutic procedures such as polypectomy, EMR, ESD, control of hemorrhage, or stricture dilation are performed. The risk of adverse events for diagnostic EUS (without needle aspiration) is similar to that for diagnostic upper endoscopy.

Infectious complications are uncommon with most endoscopic procedures. Some procedures carry a higher incidence of postprocedure bacteremia, and prophylactic antibiotics may be indicated (Table 322-1). Management of antithrombotic agents prior to endoscopic procedures should take into account the procedural risk of hemorrhage, the agent, and the patient condition, as summarized in Table 322-2.

ERCP carries additional risks. Pancreatitis occurs in ~5% of patients undergoing the procedure, and young, anicteric patients with normal ducts are at increased risk (up to 25%). Post-ERCP pancreatitis is usually mild and self-limited, but it may result in prolonged hospitalization, surgery, diabetes, or death when severe. Significant bleeding occurs after endoscopic sphincterotomy in ~1% of cases. Ascending cholangitis, pseudocyst infection, duodenal perforation, and abscess formation may occur as a result of ERCP.

Percutaneous gastrostomy tube placement during EGD is associated with a 10–15% incidence of adverse events, most often wound infections. Fascitis, pneumonia, bleeding (Fig. 322-29), buried bumper syndrome (Fig. 322-30), and colonic injury may result from gastrostomy tube placement.

URGENT ENDOSCOPY

■ ACUTE GASTROINTESTINAL HEMORRHAGE

Endoscopy is the primary diagnostic and therapeutic technique for patients with acute gastrointestinal hemorrhage. Although gastrointestinal bleeding stops spontaneously in most cases, some patients will

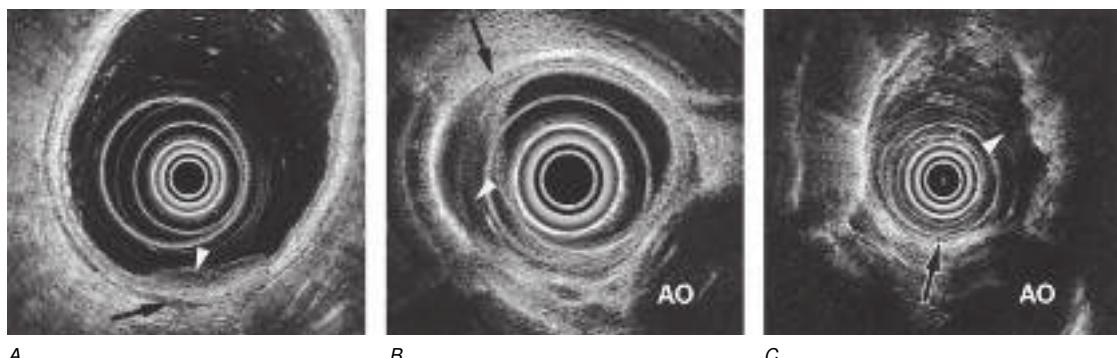


FIGURE 322-20 Local staging of gastrointestinal cancers with endoscopic ultrasound. In each example, the white arrowhead marks the primary tumor and the black arrow indicates the muscularis propria (mp) of the intestinal wall. *A*. T1 gastric cancer. The tumor does not invade the mp. *B*. T2 esophageal cancer. The tumor invades the mp. *C*. T3 esophageal cancer. The tumor extends through the mp into the surrounding tissue and focally abuts the aorta (AO).

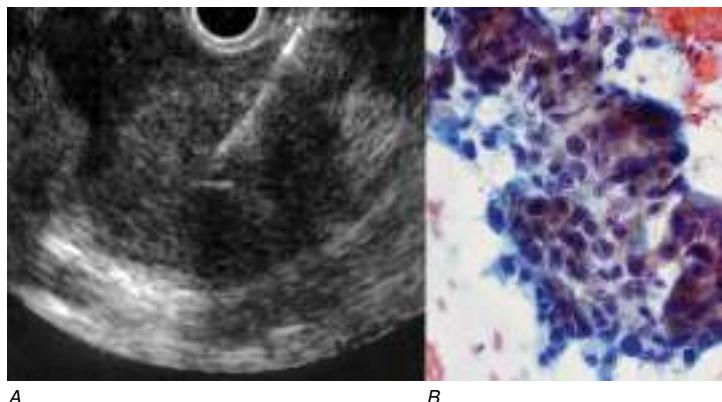


FIGURE 322-21 Endoscopic ultrasound (EUS)-guided fine-needle aspiration (FNA). *A*. Ultrasound image of a 22-gauge needle passed through the duodenal wall and positioned in a hypoechoic pancreatic head mass. *B*. Micrograph of aspirated malignant cells. (Image courtesy of Dr. Michael R. Henry.)

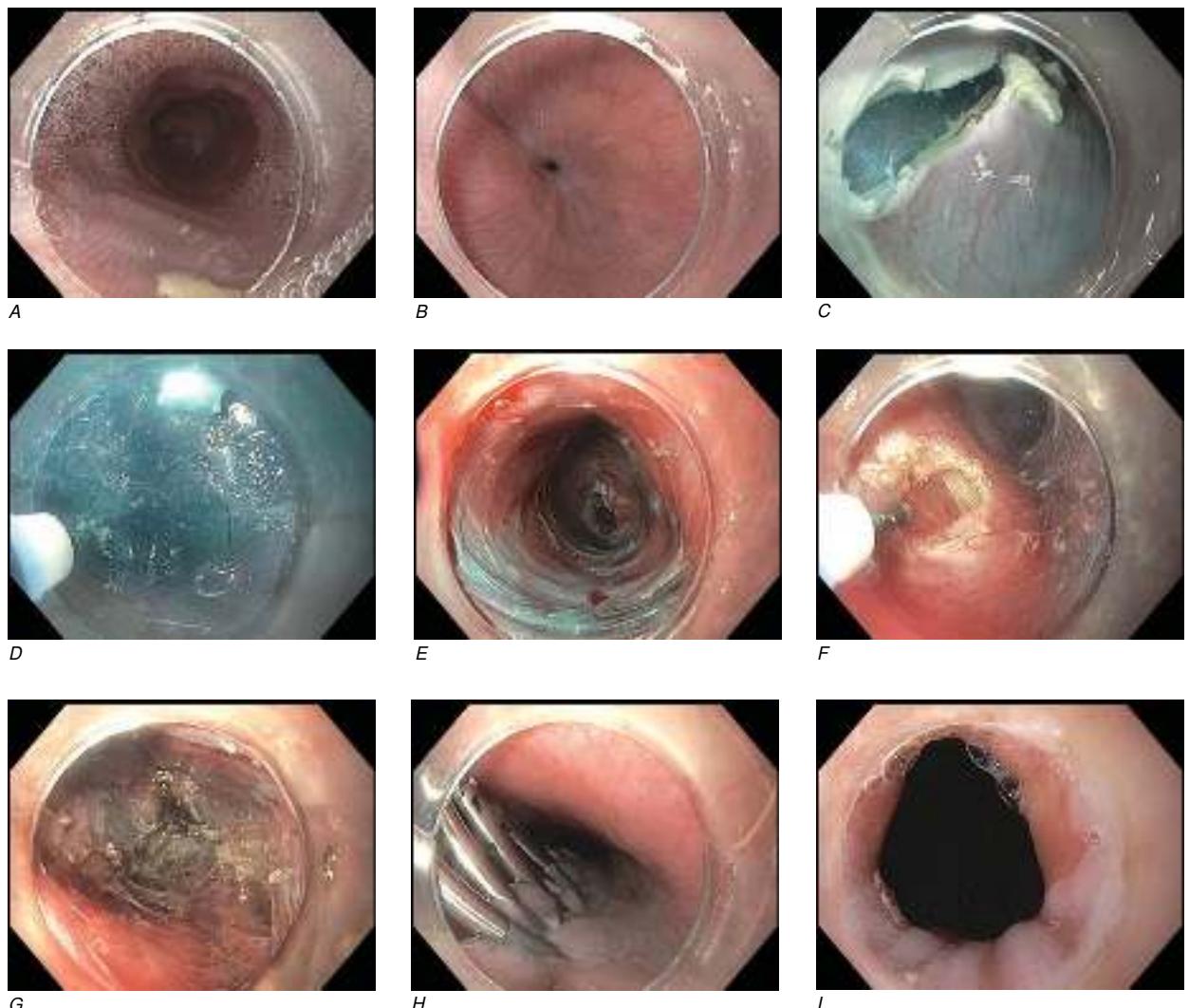


FIGURE 322-22 Peroral endoscopic myotomy (POEM) for achalasia. *A*. Dilated aperistaltic esophagus with retained secretions. *B*. Hypertonic lower esophageal sphincter (LES) region. *C*. Mucosal incision (mucosotomy) 10 cm proximal to the LES. *D*. Submucosal dissection using an electrosurgical knife following endoscope entry through the mucosotomy site into the submucosal space. *E*. Completion of submucosal tunnel to the cardia. *F*. Initiation of myotomy of the muscularis propria distal to the mucosotomy site. *G*. Completion of myotomy to the cardia. *H*. Closure of mucosotomy site with clips. *I*. Patulous gastroesophageal junction following myotomy.



FIGURE 322-23 Peroral endoscopic tumorectomy (POET). **A.** Mid-esophageal subepithelial lesion (arrow). **B.** Mucosal incision (mucosotomy) 5 cm proximal to the lesion. **C.** Submucosal dissection and tunneling to the site of the lesion. **D.** Dissection of the lesion from its attachment to the muscularis propria. **E.** Postresection defect through the muscularis propria. **F.** Mucosotomy site. **G.** Closure of mucosotomy site with clips. **H.** Resected specimen (leiomyoma).

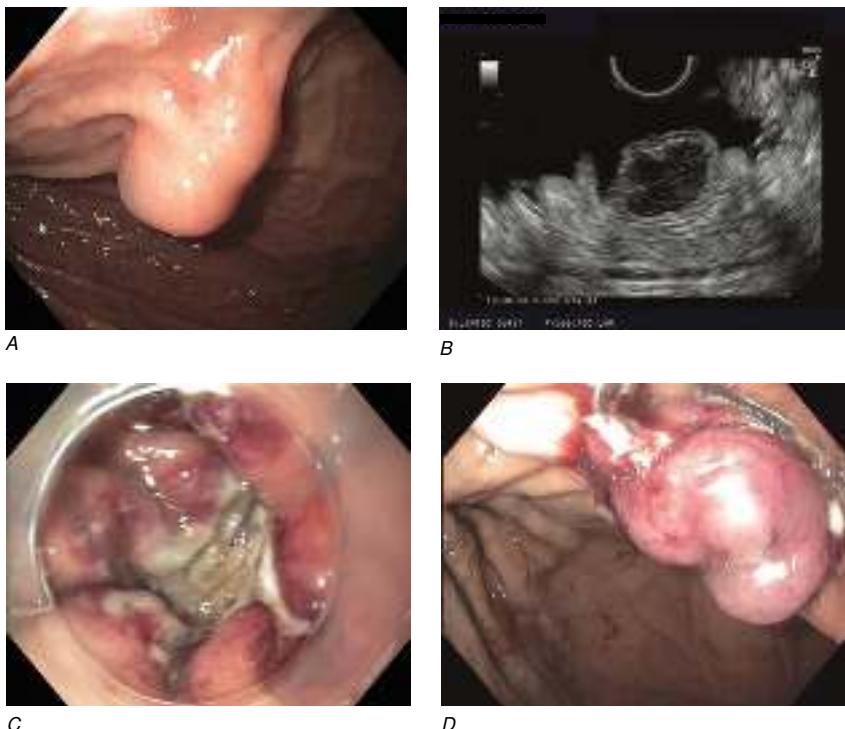


FIGURE 322-24 Endoscopic full-thickness resection (EFTR) of a gastrointestinal stromal tumor. *A*, Subepithelial lesion in the proximal stomach. *B*, Hypoechoic lesion arising from the fourth layer (muscularis propria) at endoscopic ultrasound. *C*, Full-thickness resection defect. *D*, Closure of defect using an over-the-scope clip.

have persistent or recurrent hemorrhage that may be life-threatening. Clinical predictors of rebleeding help identify patients most likely to benefit from urgent endoscopy and endoscopic, angiographic, or surgical hemostasis.

Initial Evaluation The initial evaluation of the bleeding patient focuses on the severity of hemorrhage as reflected by the presence of supine hypotension or tachycardia, postural vital sign changes, and the frequency of hematemesis or melena. Decreases in hematocrit and hemoglobin lag behind the clinical course and are not reliable gauges

of the magnitude of acute bleeding. Nasogastric tube aspiration and lavage can also be used to judge the severity of bleeding, but these are no longer routinely performed for this purpose. The bedside initial evaluation, completed well before the bleeding source is confidently identified, guides immediate supportive care of the patient, triage to the ward or intensive care unit, and timing of endoscopy. The severity of the initial hemorrhage is the most important indication for urgent endoscopy, since a large initial bleed increases the likelihood of ongoing or recurrent bleeding. Patients with resting hypotension or orthostatic change in vital signs, repeated hematemesis, or bloody nasogastric



FIGURE 322-25 Endoscopic mucosal resection (EMR). *A*, Large sessile polypoid fold in the transverse colon. *B*, Lifting of lesion following submucosal fluid injection. *C*, Piecemeal hot snare resection. *D*, Initial resection site. *E*, Resection defect following completion of piecemeal EMR.

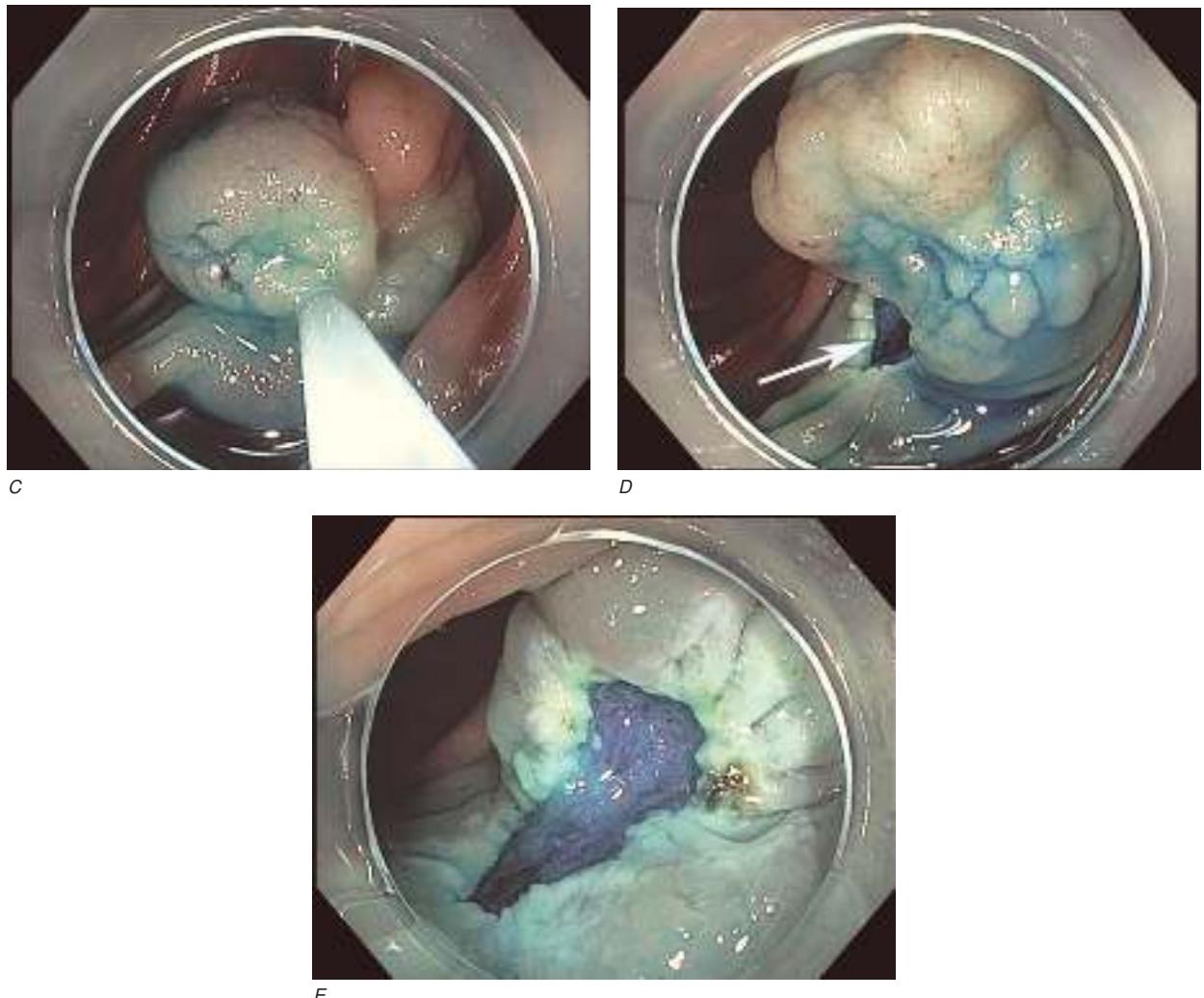


FIGURE 322-25 (Continued)

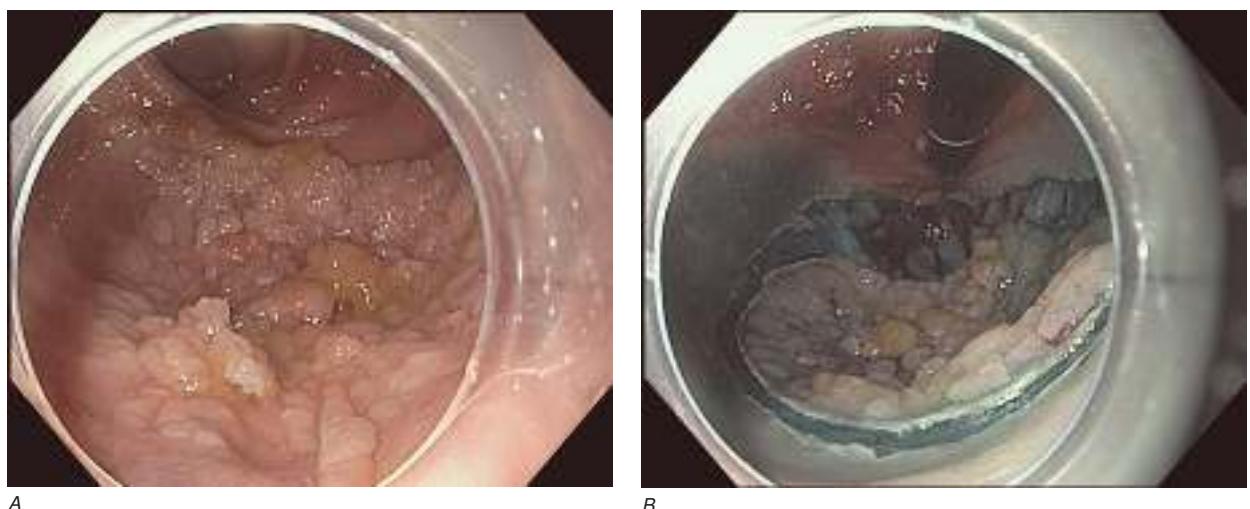


FIGURE 322-26 Endoscopic submucosal dissection (ESD). A. Large, flat, distal rectal adenoma. B. Circumferential incision following submucosal fluid injection at the periphery of the lesion. C. ESD using an electrosurgical knife. D. Rectal defect following ESD. E Specimen resected en bloc.

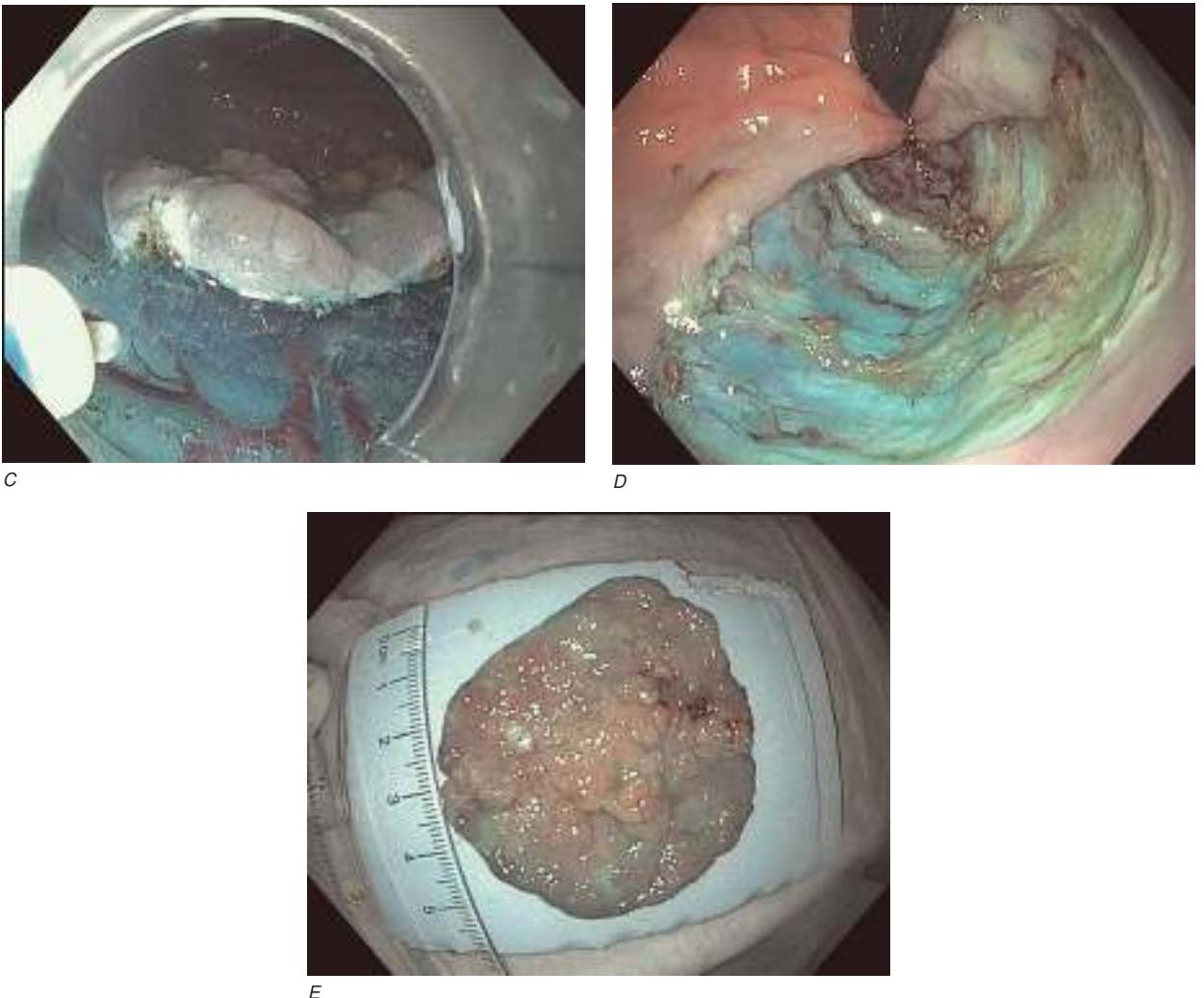


FIGURE 322-26 (Continued)

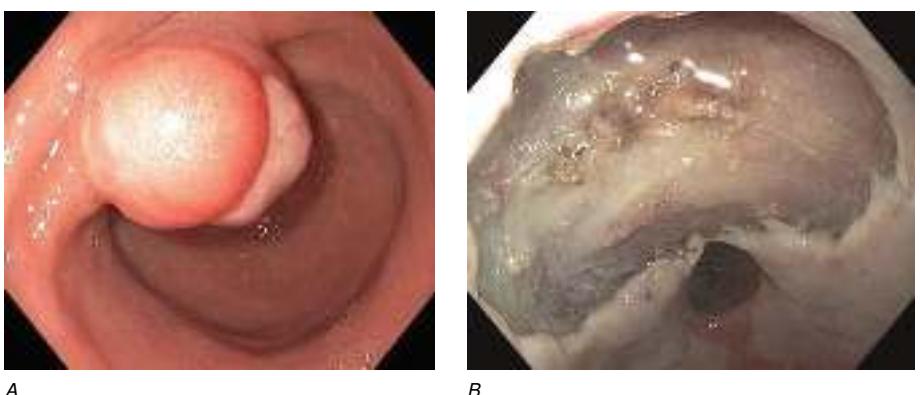


FIGURE 322-27 Closure of large defect using an endoscopic suturing device. A. Ulcerated inflammatory fibroid polyp in the antrum. B. Large defect following endoscopic submucosal dissection of the lesion. C. Closure of the defect using endoscopic sutures (arrows). D. Resected specimen.

*C**D*

FIGURE 322-27 (Continued)

aspirate that does not clear with large-volume lavage or those requiring blood transfusions should be considered for urgent endoscopy. In addition, patients with cirrhosis, coagulopathy, or respiratory or renal failure and those >70 years old are more likely to have significant rebleeding and to benefit from prompt evaluation and treatment.

Bedside evaluation also suggests an upper or lower gastrointestinal source of bleeding in most patients. Over 90% of patients with melena are bleeding proximal to the ligament of Treitz, and ~85% of patients with hematochezia are bleeding from the colon. Melena can result from bleeding in the small bowel or right colon, especially in older patients with slow colonic transit. Conversely, some patients with massive hematochezia may be bleeding from an upper gastrointestinal source,

with rapid intestinal transit. Early upper endoscopy should be considered in such patients.

Endoscopy should be performed after the patient has been resuscitated with intravenous fluids and transfusions, as necessary. Marked coagulopathy or thrombocytopenia is usually treated before endoscopy, since correction of these abnormalities may lead to resolution of bleeding, and techniques for endoscopic hemostasis are limited in such patients. Metabolic derangements should also be addressed. Tracheal intubation for airway protection should be considered before upper endoscopy in patients with repeated recent hematemesis, particularly in those with suspected variceal hemorrhage. A single dose of erythromycin (3–4 mg/kg or 250 mg) administered intravenously 30–90 min prior

*A**B**C**D*

FIGURE 322-28 Prevention of stent migration using endoscopic sutures. *A*. Esophagogastric anastomotic stricture refractory to balloon dilation. *B*. Temporary placement of a covered esophageal stent. *C*. Endoscopic suturing device to anchor the stent to the esophageal wall. *D*. Stent fixation with endoscopic sutures (arrows).

TABLE 322-1 Antibiotic Prophylaxis for Endoscopic Procedures

PATIENT CONDITION	PROCEDURE CONTEMPLATED	GOAL OF PROPHYLAXIS	PERIPROCEDURAL ANTIBIOTIC PROPHYLAXIS
All cardiac conditions	Any endoscopic procedure	Prevention of infective endocarditis	Not indicated
Bile duct obstruction in the absence of cholangitis	ERCP with complete drainage	Prevention of cholangitis	Not recommended
Bile duct obstruction in the absence of cholangitis	ERCP with anticipated incomplete drainage (e.g., sclerosing cholangitis, hilar strictures)	Prevention of cholangitis	Recommended; continue antibiotics after the procedure
Sterile pancreatic fluid collection (e.g., pseudocyst, necrosis), which communicates with pancreatic duct	ERCP	Prevention of cyst infection	Recommended; continue antibiotics after the procedure
Sterile pancreatic fluid collection	Transmural drainage	Prevention of cyst infection	Recommended
Solid lesion along upper GI tract	EUS-FNA	Prevention of local infection	Not recommended ^a
Solid lesion along lower GI tract	EUS-FNA	Prevention of local infection	Not recommended ^a
Cystic lesions along GI tract (including mediastinum and pancreas)	EUS-FNA	Prevention of cyst infection	Recommended
All patients	Percutaneous endoscopic feeding tube placement	Prevention of peristomal infection	Recommended ^b
Cirrhosis with acute GI bleeding	Required for all such patients, regardless of endoscopic procedures	Prevention of infectious complications and reduction of mortality	Recommended, upon admission ^c
Continuous peritoneal dialysis	Lower GI tract endoscopy	Prevention of bacterial peritonitis	Recommended
Synthetic vascular graft and other nonvalvular cardiovascular devices	Any endoscopic procedure	Prevention of graft and device infection	Not recommended ^d
Prosthetic joints	Any endoscopic procedure	Prevention of septic arthritis	Not recommended ^d

^aLow rates of bacteremia and local infection. ^bCefazolin or an antibiotic with equivalent coverage of oral and skin flora. ^cRisk for bacterial infection associated with cirrhosis and GI bleeding is well established; ceftriaxone or a quinolone antibiotic recommended. ^dVery low risk of infection.

Abbreviations: ERCP, endoscopic retrograde cholangiopancreatography; EUS-FNA, endoscopic ultrasound–fine-needle aspiration; GI, gastrointestinal.

Source: Reproduced with permission from MA Kashab et al: Antibiotic prophylaxis for GI endoscopy. *Gastrointest Endosc* 81:81, 2015.

to upper endoscopy increases gastric emptying and may clear blood and clots from the stomach to improve endoscopic visualization.

Most patients with hematochezia who are otherwise stable can undergo semielective colonoscopy. Controlled trials have not shown a benefit to urgent colonoscopy in patients hospitalized with hematochezia, although selected patients with massive or recurrent large-volume episodes of hematochezia should probably undergo urgent colonoscopy after a rapid colonic purge with an oral polyethylene glycol solution. Colonoscopy has a higher diagnostic yield than radionuclide bleeding scans or angiography in lower gastrointestinal bleeding, and endoscopic therapy can be applied in some cases. Urgent colonoscopy can be hindered by poor visualization due to persistent vigorous bleeding with recurrent hemodynamic instability, and other techniques (such as angiography or even emergent subtotal colectomy) must be employed. In such patients, massive bleeding originating from an upper gastrointestinal source should also be considered and excluded promptly by upper endoscopy. The anal and rectal mucosa should also be visualized endoscopically early in the course of massive rectal bleeding, as bleeding lesions in or close to the anal canal may be identified that are amenable to endoscopic or surgical transanal hemostatic techniques.

Peptic Ulcer The endoscopic appearance of peptic ulcers provides useful prognostic information and guides the need for endoscopic therapy in patients with acute hemorrhage (Fig. 322-31). A clean-based ulcer is associated with a low risk (3–5%) of rebleeding; patients with melena and a clean-based ulcer may be discharged home from the emergency room or endoscopy suite if they are young, reliable, otherwise healthy, and able to return as needed. Flat pigmented spots and adherent clots covering the ulcer base have a 10% and 20% risk of rebleeding, respectively. Flat pigmented spots do not require treatment, but endoscopic therapy is generally applied to an ulcer with an adherent clot. When a fibrin plug is seen protruding from a vessel wall in the base of an ulcer (so-called sentinel clot or visible vessel), the risk of rebleeding from the ulcer approximates 40%. This finding typically

leads to endoscopic therapy to decrease the rebleeding rate. When active spurting from an ulcer is seen, there is a 90% risk of ongoing bleeding without endoscopic or surgical therapy.

Endoscopic therapy of ulcers with high-risk stigmata typically lowers the rebleeding rate to 5–10%. Several hemostatic techniques are available, including injection of epinephrine or a sclerosant into and around the vessel (Fig. 322-32), “coaptive coagulation” of the vessel in the base of the ulcer using a thermal probe that is pressed against the site of bleeding (Fig. 322-33), placement of through-the-scope clips (Fig. 322-34) or an over-the-scope clip (Fig. 322-35), or a combination of these modalities (Video V5-8). Epinephrine injection can slow or stop active bleeding, but it is not enough as a stand-alone technique for definitive hemostasis. In conjunction with endoscopic therapy, the administration of a proton pump inhibitor decreases the risk of rebleeding and improves patient outcome.

Varices Two complementary strategies guide therapy of bleeding varices: local treatment of the bleeding varices and treatment of the underlying portal hypertension. Local therapies, including endoscopic variceal band ligation, endoscopic variceal sclerotherapy, stent placement, and balloon tamponade with a Sengstaken-Blakemore tube, effectively control acute hemorrhage in most patients, although therapies that decrease portal pressure (pharmacologic treatment, surgical shunts, or radiologically placed intrahepatic portosystemic shunts) also play an important role.

Endoscopic variceal ligation (EVL) is indicated for the prevention of a first bleed (primary prophylaxis) from large esophageal varices (Fig. 322-36), particularly in patients in whom nonselective beta blockers are contraindicated or not tolerated. EVL is also the preferred endoscopic therapy for control of active esophageal variceal bleeding and for subsequent eradication of esophageal varices (secondary prophylaxis). During EVL, a varix is suctioned into a cap fitted on the end of the endoscope, and a rubber band is released from the cap, ligating the varix (Fig. 322-37, Video V5-9). EVL controls acute hemorrhage

DRUG	BLEEDING RISK OF PROCEDURE	MANAGEMENT	INTERVAL BETWEEN LAST DOSE AND PROCEDURE	COMMENTS
Warfarin	Low ^a High ^b	Continue Discontinue	N/A 3–7 days (usually 5), INR should be ≤1.5 for procedure	Ensure that INR is not supratherapeutic Consider bridging therapy with heparin ^c ; usually safe to resume warfarin on the same or next day For life-threatening GI hemorrhage, consider reversal with unactivated prothrombin complex concentrate
Dabigatran, rivaroxaban, apixaban, edoxaban	Low ^a	Continue or hold morning dose on day of procedure	N/A	
Dabigatran	High ^b	Discontinue	2–3 days if GFR is ≥50 mL/min, 3–4 days if GFR is 30–49 mL/min	Bridging therapy not recommended; resume drug when bleeding risk is low For life-threatening GI hemorrhage, consider use of a reversal agent
Rivaroxaban, apixaban, edoxaban	High ^a	Discontinue	2 days if GFR is ≥60 mL/min, 3 days if GFR is 30–59 mL/min, 4 days if GFR is <30 mL/min	Bridging therapy not recommended; resume drug when bleeding risk is low For life-threatening GI hemorrhage, consider use of a reversal agent
Heparin	Low ^a High ^b	Continue Discontinue	N/A 4–6 h for unfractionated heparin	Skip one dose if using low-molecular-weight heparin
Aspirin	Any	Continue	N/A	Low-dose aspirin does not substantially increase the risk of endoscopic procedures
Aspirin with dipyridamole	Low ^a High ^b	Continue Discontinue	N/A 2–7 days	Consider continuing aspirin monotherapy
P2Y ₁₂ receptor antagonists (clopidogrel, prasugrel, ticlopidine, ticagrelor, cangrelor)	Low ^a High ^b	Continue Coronary stent in place: discuss with cardiologist No coronary stent: discontinue, consider substituting aspirin	N/A 5 days (clopidogrel or ticagrelor), 7 days (prasugrel), 10–14 days (ticlopidine)	Risk of stent thrombosis for at least 12 months after insertion of drug-eluting coronary stent or 1 month after insertion of bare metal coronary stent

^aLow-risk endoscopic procedures include esophagogastroduodenoscopy (EGD) or colonoscopy with or without biopsy, endoscopic ultrasound (EUS) without fine-needle aspiration (FNA), and endoscopic retrograde cholangiopancreatography (ERCP) with stent exchange. ^bHigh-risk endoscopic procedures include EGD or colonoscopy with dilation, polypectomy, or thermal ablation; percutaneous endoscopic gastrostomy (PEG); EUS with FNA; and ERCP with sphincterotomy or pseudocyst drainage. ^cBridging therapy with low-molecular-weight heparin should be considered for patients discontinuing warfarin who are at high risk for thromboembolism, including those with (1) atrial fibrillation with a CHA₂DS₂-VASc score ≥3, mechanical valve(s), or history of stroke or transient ischemic attack; (2) mechanical mitral valve; (3) mechanical aortic valve with other thromboembolic risk factors or older-generation mechanical aortic valve; or (4) venous thromboembolism within the past 3 months.

Abbreviations: GFR, glomerular filtration rate; INR, international normalized ratio; N/A, not applicable.

Source: Adapted from RD Acosta et al: Gastrointest Endosc 83:3, 2016; and AM Veitch et al: Gut 65:374, 2016.

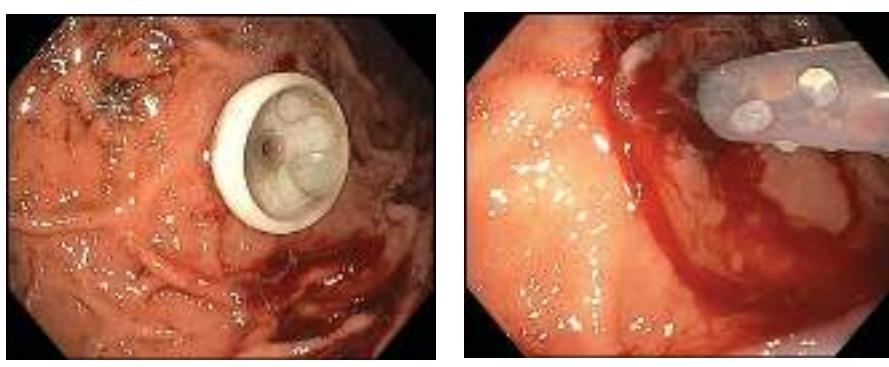


FIGURE 322-29 Bleeding from percutaneous endoscopic gastrostomy (PEG) tube placement. A. Patient with melena from a recently placed PEG tube. B. Loosening of the internal disk bumper of the PEG tube revealed active bleeding from within the PEG tract.

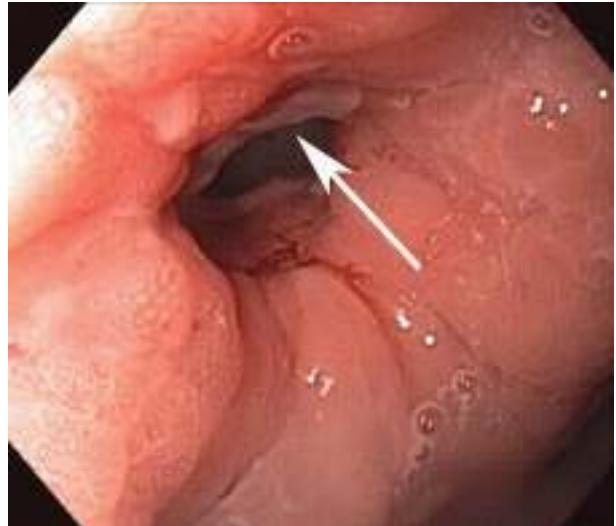
**A****B**

FIGURE 322-30 Buried bumper syndrome. **A.** Migration of the internal disk bumper of a percutaneous endoscopic gastrostomy (PEG) tube through the gastric wall. **B.** Close-up view of the disk bumper (arrow) buried in the gastric wall.

in up to 90% of patients. Complications of EVL, such as postligation ulcer bleeding and esophageal stenosis, are uncommon. Endoscopic variceal sclerotherapy (EVS) involves the injection of a sclerosing, thrombogenic solution into or next to esophageal varices. EVS also controls acute hemorrhage in most patients, but it is generally used as salvage therapy when band ligation fails because of its higher complication rate. Bleeding from large gastric fundal varices (Fig. 322-38) is best treated with endoscopic cyanoacrylate (“glue”) injection (Video V5-10), since EVL or EVS of these varices is associated with a high rebleeding rate. Complications of cyanoacrylate injection include infection and glue embolization to other organs, such as the lungs, brain, and spleen.

After treatment of the acute hemorrhage, an elective course of endoscopic therapy can be undertaken with the goal of eradicating esophageal varices and preventing rebleeding months to years later. However, this chronic therapy is less successful, preventing long-term rebleeding in ~50% of patients. Pharmacologic therapies that decrease

portal pressure have similar efficacy. The preferred strategy, however, for secondary prophylaxis of variceal bleeding is the combination of EVL with a nonselective beta blocker.

Dieulafoy's Lesion This lesion, also called persistent caliber artery, is a large-caliber arteriole that runs immediately beneath the gastrointestinal mucosa and bleeds through a focal mucosal erosion (Fig. 322-39). Dieulafoy's lesion commonly involves the lesser curvature of the proximal stomach, causes impressive arterial hemorrhage, and may be difficult to diagnose when not actively bleeding; it is often recognized only after repeated endoscopy for recurrent bleeding. Endoscopic therapy, such as thermal coagulation, band ligation, clip placement, or endoscopic suturing, is typically effective for control of bleeding and sealing of the underlying vessel once the lesion has been identified (Video V5-11). Rescue therapies, such as angiographic embolization or surgical oversewing, are considered in situations where endoscopic therapy has failed.

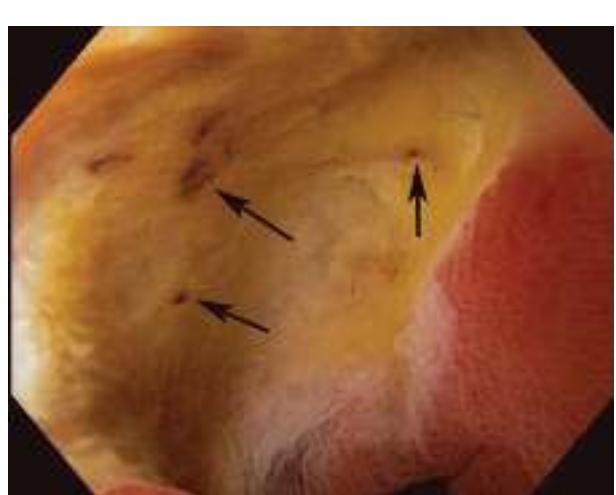
**A****B**

FIGURE 322-31 Stigmata of hemorrhage in peptic ulcers. **A.** Gastric antral ulcer with a clean base. **B.** Duodenal ulcer with flat pigmented spots (arrows). **C.** Duodenal ulcer with a dense adherent clot. **D.** Duodenal ulcer with a pigmented protuberance/visible vessel. **E.** Duodenal ulcer with active spurting (arrow).

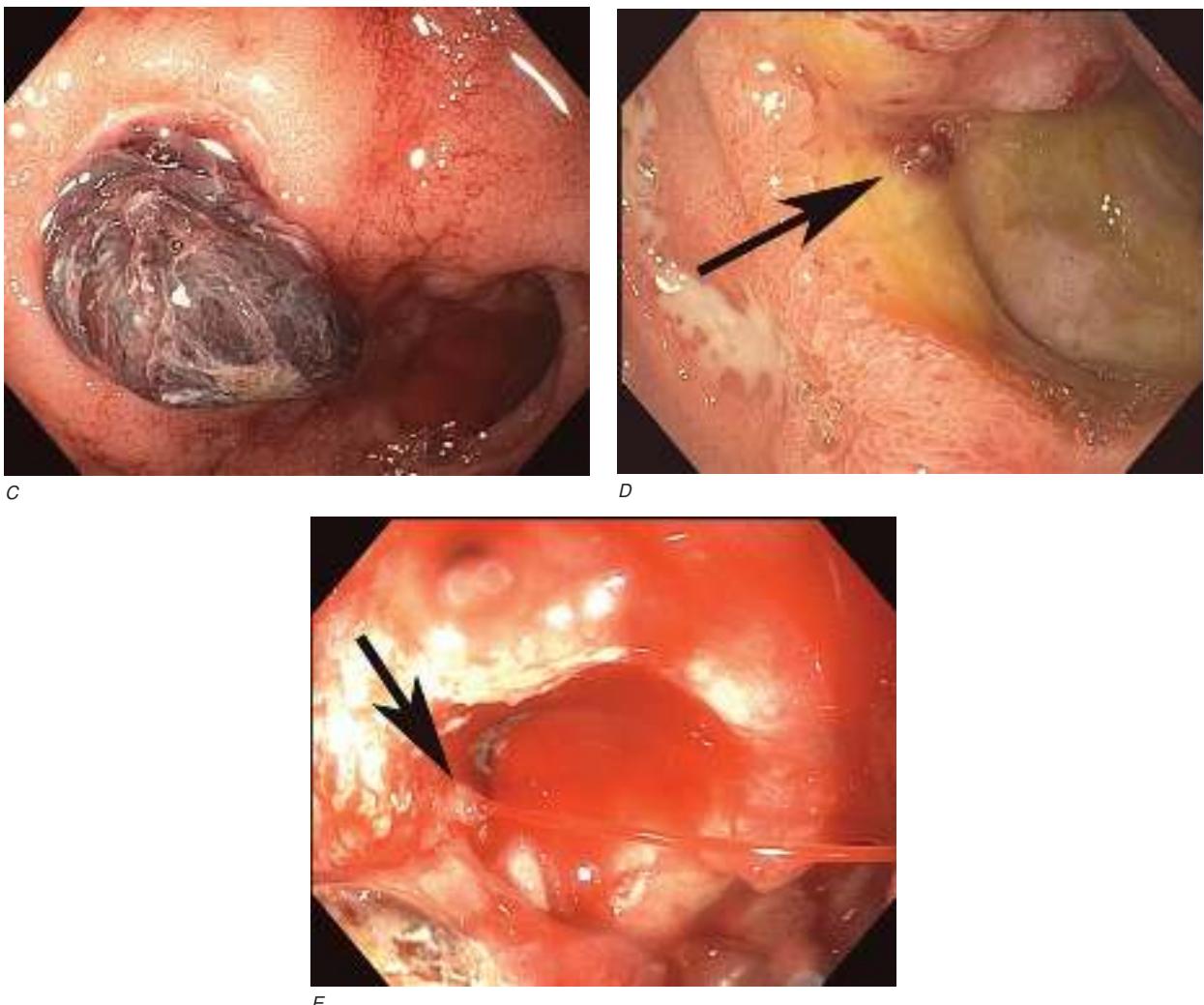


FIGURE 322-31 (Continued)



FIGURE 322-32 Injection therapy for ulcer hemostasis. Epinephrine injection into a duodenal ulcer with visible vessel (arrow) and adherent clot.

Mallory-Weiss Tear A Mallory-Weiss tear is a linear mucosal rent near or across the gastroesophageal junction that is often associated with retching or vomiting (Fig. 322-40). When the tear disrupts a submucosal arteriole, brisk hemorrhage may result. Endoscopy is the best method for diagnosis, and an actively bleeding tear can be treated endoscopically with coaptive coagulation, band ligation, or hemoclips, with or without epinephrine injection (Video V5-12). Unlike peptic ulcer, a Mallory-Weiss tear with a nonbleeding sentinel clot in its base rarely rebleeds and thus does not necessitate endoscopic therapy.

Vascular Ectasias Vascular ectasias are flat mucosal vascular anomalies that are best diagnosed by endoscopy. They usually cause slow intestinal blood loss and occur either in a sporadic fashion or in a well-defined pattern of distribution (e.g., gastric antral vascular ectasia [GAVE] or “watermelon stomach”) (Fig. 322-41). Cecal vascular ectasias, GAVE, and radiation-induced rectal ectasias are often responsive to local endoscopic ablative therapy, such as argon plasma coagulation (Video V5-13). Patients with diffuse small-bowel vascular ectasias (associated with chronic renal failure and with hereditary hemorrhagic telangiectasia) may continue to bleed despite endoscopic treatment of easily accessible lesions by conventional endoscopy. These patients may benefit from device-assisted enteroscopy with endoscopic hemostasis or pharmacologic therapy, such as octreotide or low-dose thalidomide, in those who continue to bleed despite endoscopic therapy.

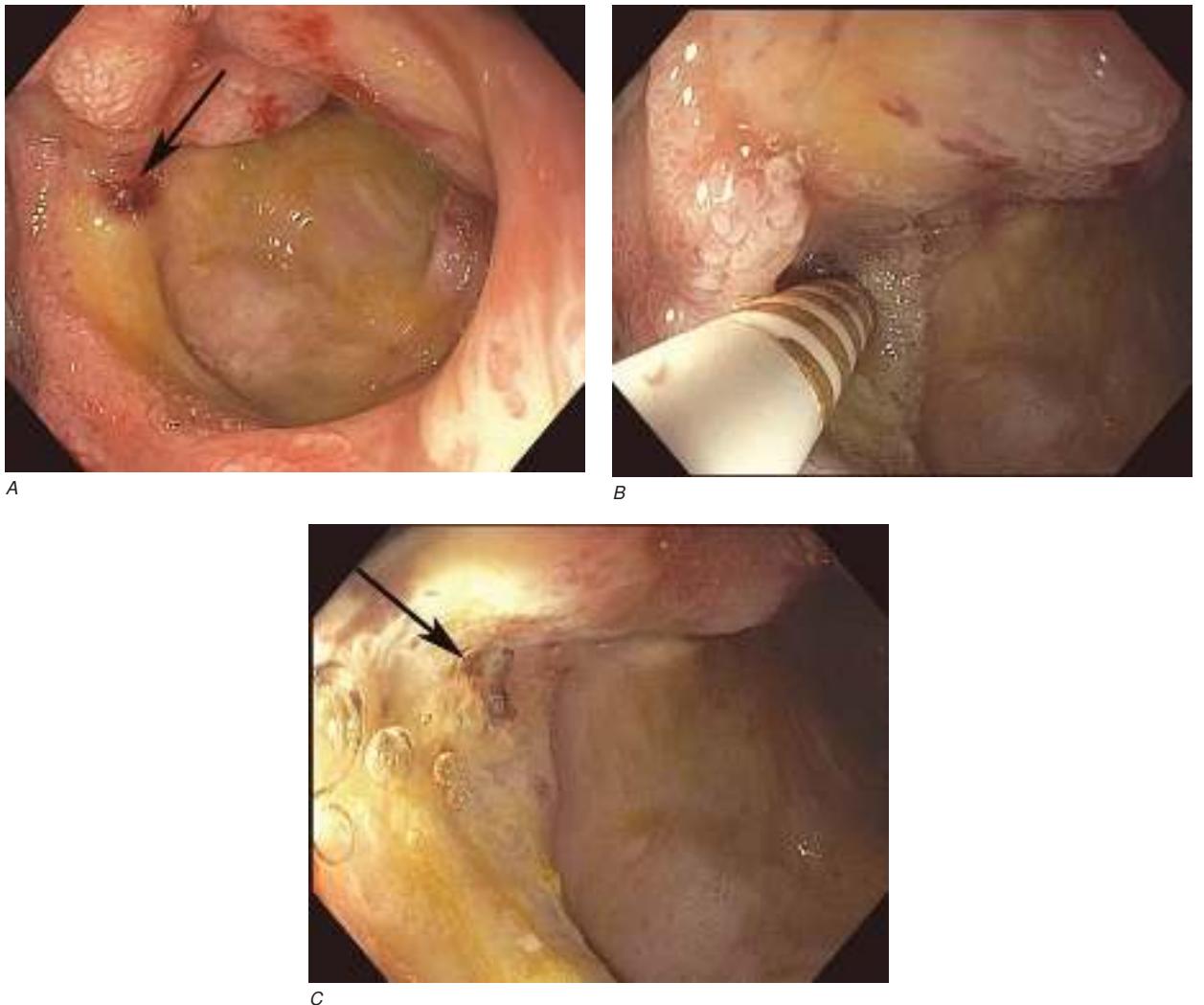


FIGURE 322-33 Contact coagulation for ulcer hemostasis. A. Duodenal ulcer with a visible vessel (arrow). B. Coagulation of the vessel with a contact thermal probe. C. Obliteration of the treated vessel (arrow).



FIGURE 322-34 Through-the-scope clip placement for ulcer hemostasis. A. Superficial duodenal ulcer with visible vessel (arrow). B. Hemostasis secured following placement of multiple through-the-scope clips.

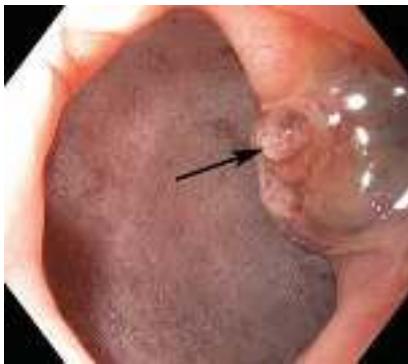
*A**B*

FIGURE 322-35 Over-the-scope clip placement for ulcer hemostasis. *A*, Pyloric channel ulcer with visible vessel (arrow). *B*, Hemostasis secured following placement of an over-the-scope clip.

Colonic Diverticula Diverticula form where nutrient arteries penetrate the muscular wall of the colon en route to the colonic mucosa ([Fig. 322-42](#)). The artery found in the base of a diverticulum may bleed, causing painless and impressive hematochezia. Colonoscopy is indicated in patients with hematochezia and suspected diverticular hemorrhage, since other causes of bleeding (such as vascular ectasias, colitis, and colon cancer) must be excluded. In addition, an actively bleeding diverticulum may be seen and treated during colonoscopy ([Fig. 322-43](#), [Video V5-14](#)).

GASTROINTESTINAL OBSTRUCTION AND PSEUDOObSTRUCTION

Endoscopy is useful for evaluation and treatment of some forms of gastrointestinal obstruction. An important exception is small-bowel obstruction due to surgical adhesions, which is generally not diagnosed



FIGURE 322-36 Esophageal varices.

*A**B*

FIGURE 322-37 Endoscopic variceal ligation. *A*, Esophageal varices with red wale marks. *B*, Band ligation of varices.

or treated endoscopically. Esophageal, gastroduodenal, and colonic obstruction or pseudoobstruction can all be diagnosed and often managed endoscopically.

Acute Esophageal Obstruction Esophageal obstruction by impacted food ([Fig. 322-44](#)) or an ingested foreign body ([Fig. 322-45](#)) is a potentially life-threatening event and represents an endoscopic emergency. Left untreated, the patient may develop esophageal ulceration, ischemia, and perforation. Patients with persistent esophageal obstruction often have hypersalivation and are usually unable to swallow water. Sips of a carbonated beverage, sublingual nifedipine or nitrates, or intravenous glucagon may resolve an esophageal food impaction, but in many patients, an underlying web, ring, or stricture is present, and endoscopic removal of the obstructing food bolus is necessary. Endoscopy is generally the best initial test in such patients since endoscopic removal of the obstructing material is usually possible, and the presence of an underlying esophageal pathology can often be determined. Radiographs of the chest and neck should be considered before endoscopy in patients with fever, obstruction for ≥ 24 h, or ingestion of a sharp object, such as a fishbone. Radiographic contrast studies interfere with subsequent endoscopy and are not advisable in most patients with a clinical picture of esophageal obstruction.



FIGURE 322-38 Gastric varices. A. Large gastric fundal varices. B. Stigmata of recent bleeding from the same gastric varices (arrow).

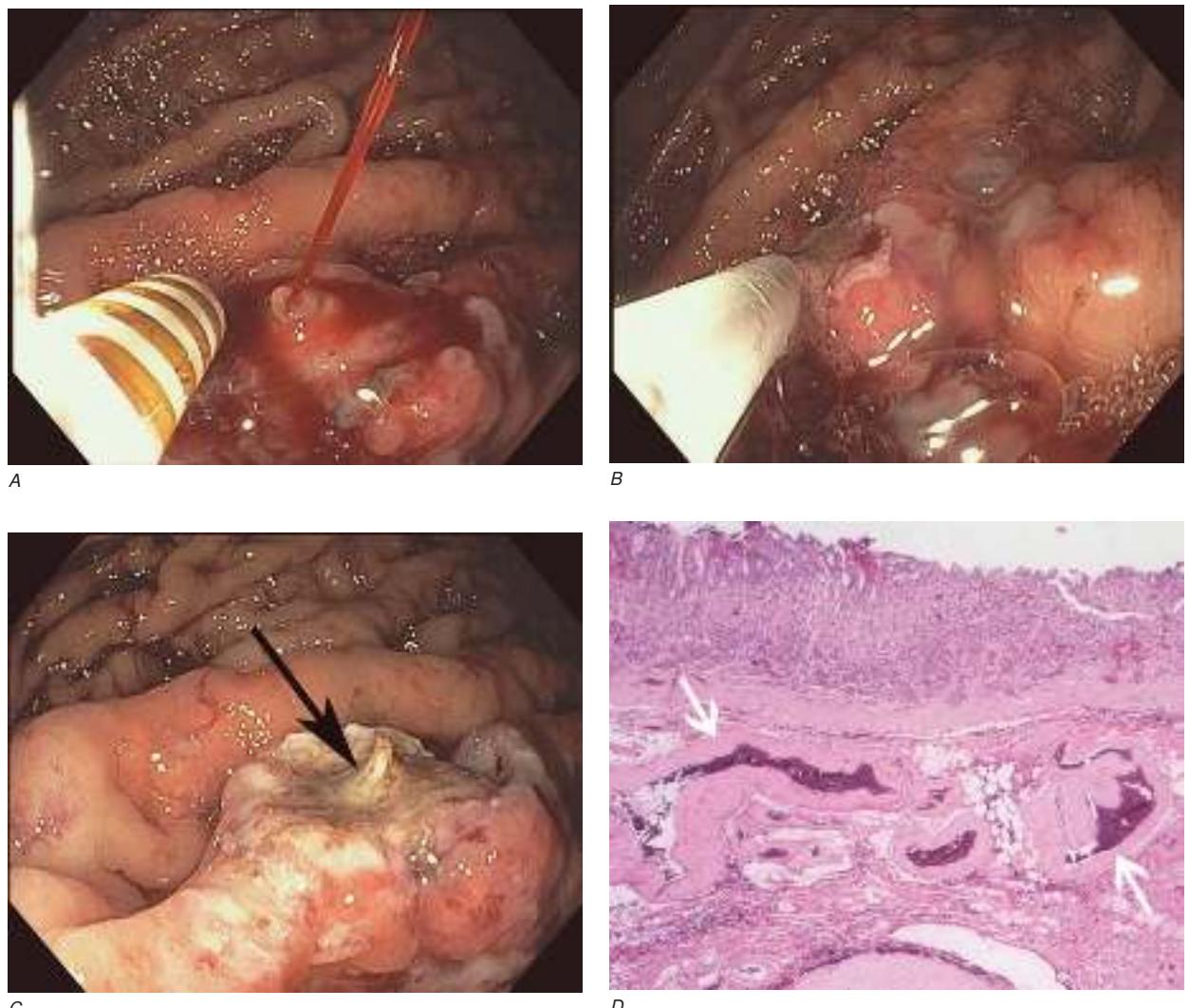


FIGURE 322-39 Dieulafoy's lesion. A. Actively spouting gastric Dieulafoy's lesion. B. Coagulation of the lesion using a contact thermal probe. C. Hemostasis secured following contact coagulation (arrow). D. Histology of a gastric Dieulafoy's lesion. A persistent caliber artery (arrows) is present in the gastric submucosa, immediately beneath the mucosa.

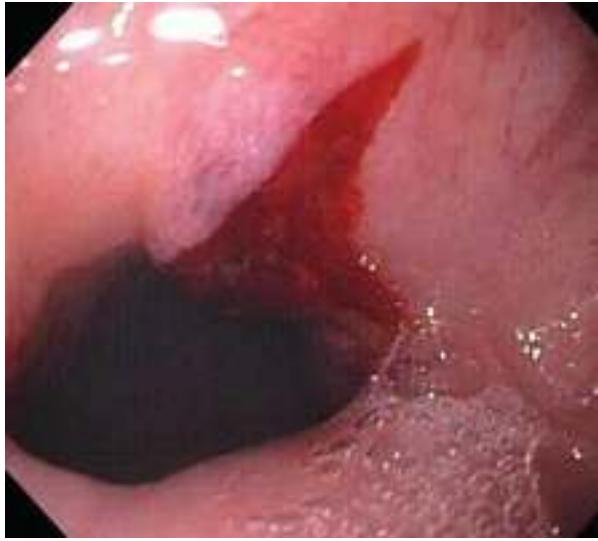


FIGURE 322-40 Mallory-Weiss tear at the gastroesophageal junction.

Gastric Outlet Obstruction Obstruction of the gastric outlet is commonly caused by gastric, duodenal, or pancreatic malignancy or chronic peptic ulceration with stenosis of the pylorus (Fig. 322-46). Patients vomit partially digested food many hours after eating. Gastric decompression with a nasogastric tube and subsequent lavage for removal of retained material is the first step in treatment. Endoscopy is useful for diagnosis and treatment. Patients with benign pyloric stenosis may be treated with endoscopic balloon dilation of the pylorus, and a course of endoscopic dilation results in long-term relief of symptoms in ~50% of patients. Removable, fully covered lumen-apposing metal stents (LAMS) may also be used to treat benign pyloric stenosis (Video V5-15). Malignant gastric outlet obstruction can be relieved with endoscopically placed expandable stents in patients with inoperable malignancy (Video V5-16).

Colonic Obstruction and Pseudoobstruction These conditions both present with abdominal distention and discomfort, tympany, and a dilated colon on plain abdominal radiography. The radiographic appearance may be characteristic of a particular condition, such as sigmoid volvulus (Fig. 322-47). Both obstruction and pseudoobstruction may lead to colonic perforation if left untreated. Acute colonic pseudoobstruction is a form of colonic ileus that is usually attributable to electrolyte disorders, narcotic and anticholinergic medications, immobility (as after surgery), or retroperitoneal hemorrhage or mass. Multiple causative factors are often present. Colonoscopy, water-soluble contrast enema, or CT may be used to assess for an obstructing lesion and differentiate obstruction from pseudoobstruction. One of these diagnostic studies should be strongly considered if the patient does not have clear risk factors for pseudoobstruction, if radiographs do not show air in the rectum, or if the patient fails to improve when underlying causes of pseudoobstruction have been addressed. The risk of cecal perforation in pseudoobstruction rises when the cecal diameter exceeds 12 cm, and decompression of the colon may be achieved using intravenous neostigmine or via colonoscopic decompression (Fig. 322-48). Most patients should receive a trial of conservative therapy (with correction of electrolyte disorders, removal of offending medications, and increased mobilization) before undergoing an invasive decompressive procedure for colonic pseudoobstruction.

Colonic obstruction is an indication for urgent intervention. In the past, emergent diverting colostomy was usually performed with a subsequent second operation after bowel preparation to treat the underlying cause of obstruction. Colonoscopic placement of an expandable stent is an alternative treatment option that can relieve malignant colonic obstruction without emergency surgery and permit

bowel preparation for an elective one-stage operation (Fig. 322-49; Video V5-17).

■ ACUTE BILIARY OBSTRUCTION

The steady, severe pain that occurs when a gallstone acutely obstructs the common bile duct often brings patients to a hospital. The diagnosis of a ductal stone is suspected when the patient is jaundiced or when serum liver tests or pancreatic enzyme levels are elevated; it is confirmed by EUS, magnetic resonance cholangiopancreatography (MRCP), or direct cholangiography (performed endoscopically, percutaneously, or during surgery). ERCP is the primary means of treating common bile duct stones (Figs. 322-16 and 322-17), although they can also be removed by bile duct exploration at the time of cholecystectomy. Radiologic percutaneous biliary drainage may be required in some cases.

Bile Duct Imaging While transabdominal ultrasound diagnoses only a minority of bile duct stones, MRCP and EUS are >90% accurate and have an important role in diagnosis. Examples of these modalities are shown in Fig. 322-50.

If the suspicion for a bile duct stone is high and urgent treatment is required (as in a patient with obstructive jaundice and biliary sepsis), ERCP is the procedure of choice since it remains the gold standard for diagnosis and allows for immediate treatment (Video V5-18). If a persistent bile duct stone is relatively unlikely (as in a patient with gallstone pancreatitis), ERCP may be supplanted by less invasive imaging techniques, such as EUS, MRCP, or intraoperative cholangiography performed during cholecystectomy, sparing some patients the risk and discomfort of ERCP.

Ascending Cholangitis Charcot's triad of jaundice, abdominal pain, and fever is present in ~70% of patients with ascending cholangitis and biliary sepsis. These patients are managed initially with fluid resuscitation and intravenous antibiotics. Abdominal ultrasound is often performed to assess for gallbladder stones and bile duct dilation. However, the bile duct may not be dilated early in the course of acute biliary obstruction. Medical management usually improves the patient's clinical status, providing a window of ~24 h during which biliary drainage should be established, typically by ERCP. Undue delay can result in recrudescence of overt sepsis and increased morbidity and mortality rates. In addition to Charcot's triad, the additional presence of shock and confusion (Reynolds's pentad) is associated with a high mortality rate and should prompt urgent intervention to restore biliary drainage.

Gallstone Pancreatitis Gallstones may cause acute pancreatitis as they pass through the ampulla of Vater. The occurrence of gallstone pancreatitis usually implies passage of a stone into the duodenum, and only ~20% of patients harbor a persistent stone in the ampulla or the common bile duct. Retained stones are more common in patients with jaundice, rising serum liver tests following hospitalization, severe pancreatitis, or superimposed ascending cholangitis.

Urgent ERCP decreases the morbidity rate of gallstone pancreatitis in a subset of patients with retained bile duct stones. It is unclear whether the benefit of ERCP is mainly attributable to treatment and prevention of ascending cholangitis or to relief of pancreatic ductal obstruction. ERCP is warranted early in the course of gallstone pancreatitis if ascending cholangitis is suspected, especially in a jaundiced patient. Urgent ERCP may also benefit patients predicted to have severe pancreatitis using a clinical index of severity, such as the Glasgow or Ranson score. Since the benefit of ERCP is limited to patients with a retained bile duct stone, a strategy of initial MRCP or EUS for diagnosis decreases the utilization of ERCP in gallstone pancreatitis and improves clinical outcomes by limiting the occurrence of ERCP-related adverse events.

ELECTIVE ENDOSCOPY

■ DYSPEPSIA

Dyspepsia is a chronic or recurrent burning discomfort or pain in the upper abdomen that may be caused by diverse processes, such as gastroesophageal reflux, peptic ulcer disease, and "nonulcer dyspepsia,"

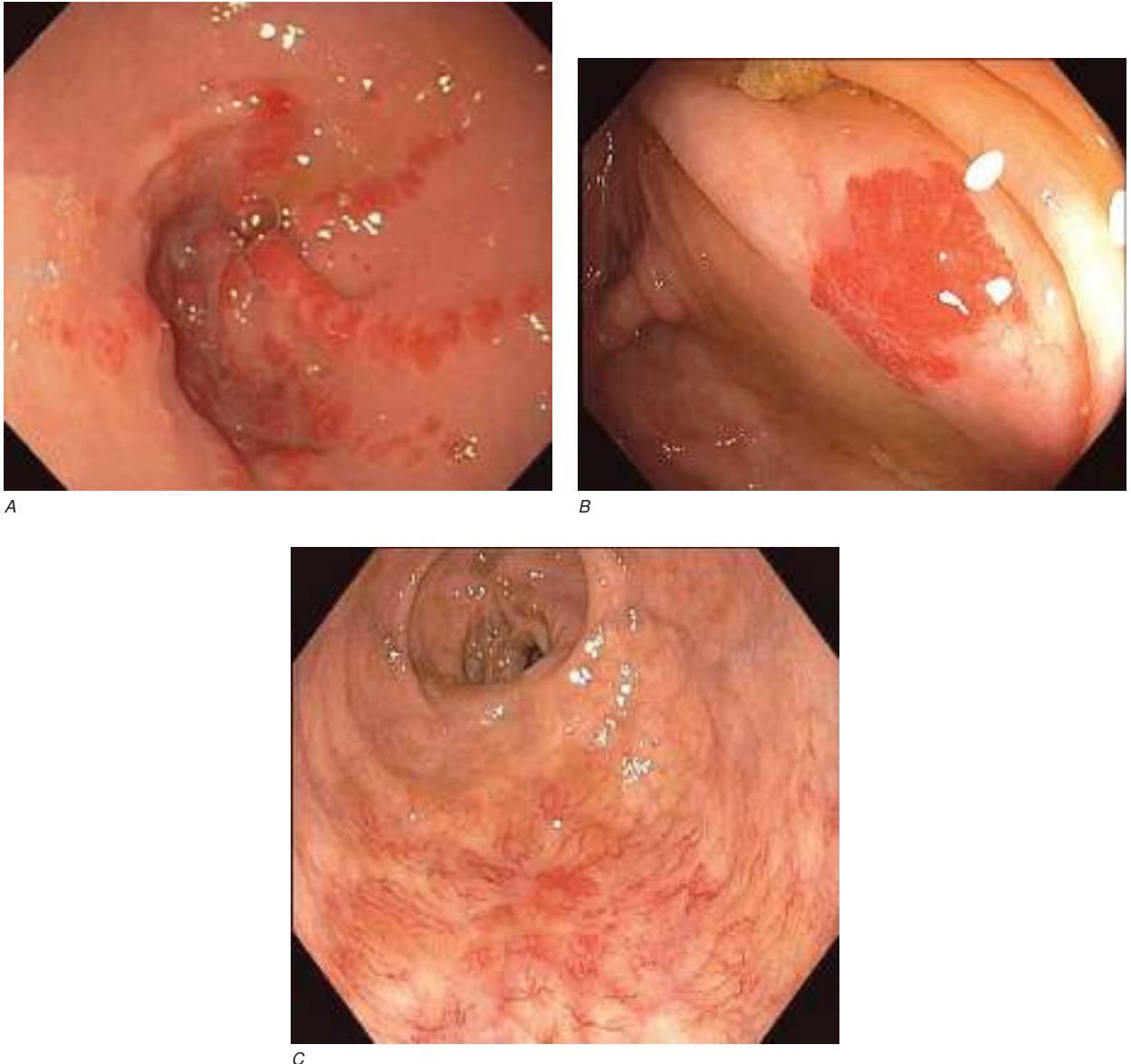


FIGURE 322-41 Gastrointestinal vascular ectasias. *A*. Gastric antral vascular ectasia (“watermelon stomach”) characterized by stripes of prominent flat or raised vascular ectasias. *B*. Cecal vascular ectasia. *C*. Radiation-induced vascular ectasias of the rectum in a patient previously treated for prostate cancer.

a heterogeneous category that includes disorders of motility, sensation, and somatization. Gastric and esophageal malignancies are less common causes of dyspepsia. Careful history-taking allows accurate differential diagnosis of dyspepsia in only about half of patients. In the remainder, endoscopy can be a useful diagnostic tool, especially in patients whose symptoms are not resolved by *Helicobacter pylori* treatment or an empirical trial of acid-reducing therapy. Endoscopy should be performed at the outset in patients with dyspepsia and alarm features, such as weight loss, obstructive symptoms, or iron-deficiency anemia.

■ GASTROESOPHAGEAL REFLUX DISEASE

When classic symptoms of gastroesophageal reflux are present, such as water brash and substernal heartburn, presumptive diagnosis and empirical treatment are often sufficient. Endoscopy is a sensitive test for diagnosis of esophagitis (Fig. 322-51), but it will miss nonerosive reflux disease (NERD) since some patients have symptomatic

reflux without esophagitis. The most sensitive test for diagnosis of gastroesophageal reflux disease (GERD) is 24-h ambulatory pH monitoring. Endoscopy is indicated in patients with reflux symptoms refractory to antisecretory therapy; in those with alarm symptoms, such as dysphagia, weight loss, or gastrointestinal bleeding; and in those with recurrent dyspepsia after treatment that is not clearly due to reflux on clinical grounds alone. Endoscopy should be considered in patients with long-standing (≥ 10 years) GERD, as they have a sixfold increased risk of harboring Barrett's esophagus compared to patients with <1 year of reflux symptoms.

Barrett's Esophagus and Esophageal Squamous Dysplasia Barrett's esophagus is specialized columnar metaplasia that replaces the normal squamous mucosa of the distal esophagus in some persons with GERD. Barrett's epithelium is a major risk factor for adenocarcinoma of the esophagus and is readily detected endoscopically, due to proximal displacement of the squamocolumnar junction (Fig. 322-6).



FIGURE 322-42 Colonic diverticula.

A screening EGD for Barrett's esophagus should be considered in patients with a chronic (≥ 10 year) history of GERD symptoms. Endoscopic biopsy is the gold standard for confirmation of Barrett's esophagus and for dysplasia or cancer arising in Barrett's mucosa.

Periodic EGD with biopsies is recommended for surveillance of patients with Barrett's esophagus. Endoscopic resection (EMR or ESD) and/or ablation are performed when high-grade dysplasia or intramucosal cancer are found in the Barrett's mucosa. Both endoscopic therapy and periodic surveillance are acceptable options in patients with Barrett's esophagus and low-grade dysplasia. Radiofrequency ablation (RFA) is the most common ablative modality used for endoscopic treatment of Barrett's esophagus, and other modalities, such as cryotherapy, are also available.

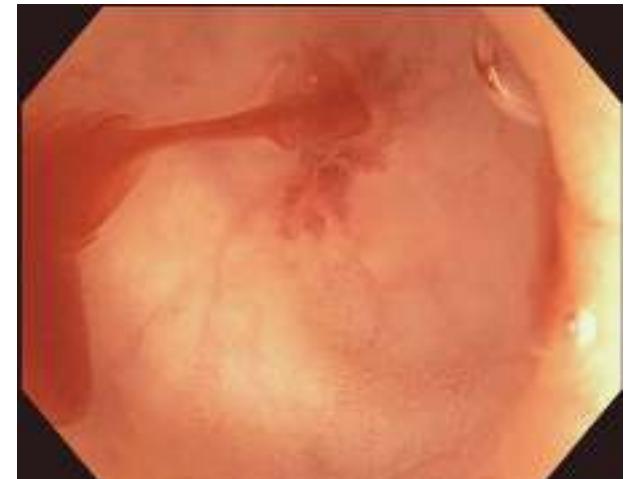
Esophageal squamous dysplasia is the precursor lesion of esophageal squamous cell cancer (ESCC), the most common type of esophageal malignancy worldwide. Endoscopic detection of esophageal squamous dysplasia often requires specialized imaging methods, such as chromoendoscopy with Lugol's iodine. Once detected, it can be treated endoscopically with EMR, ESD, or RFA (Fig. 322-52). Population-based screening for esophageal squamous dysplasia has been shown to decrease the occurrence of ESCC in high-incidence regions.

■ PEPTIC ULCER

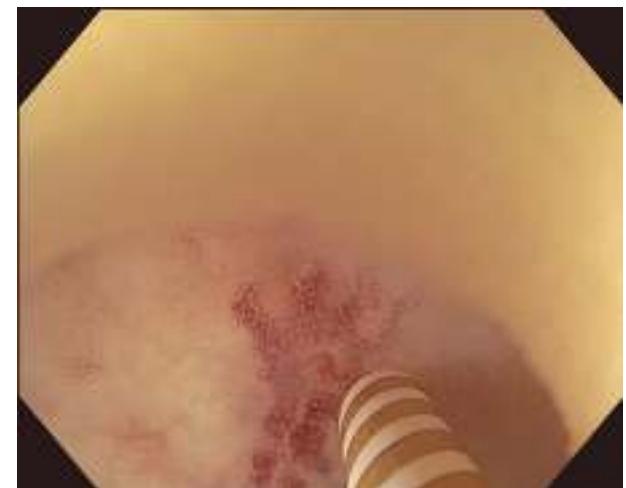
Peptic ulcer classically causes epigastric gnawing or burning, often occurring nocturnally and promptly relieved by food or antacids. Although endoscopy is the most sensitive diagnostic test for peptic ulcer, it is not a cost-effective strategy in young patients with ulcer-like dyspeptic symptoms unless endoscopy is available at low cost. Patients with suspected peptic ulcer should be evaluated for *H. pylori* infection. Serology (past or present infection), urea breath testing (current infection), and stool tests are noninvasive and less costly than endoscopy with biopsy. Patients aged >50 and those with alarm symptoms or persistent symptoms despite treatment should undergo endoscopy to exclude malignancy.

■ NONULCER DYSPEPSIA

Nonulcer dyspepsia may be associated with bloating and, unlike peptic ulcer, tends not to remit and recur. Most patients describe persistent symptoms despite acid-reducing, prokinetic, or anti-*Helicobacter* therapy and are referred for endoscopy to exclude a refractory ulcer and assess for other causes. Although endoscopy is useful for excluding other diagnoses, its impact on the treatment of patients with nonulcer dyspepsia is limited.



A



B



C

FIGURE 322-43 Diverticular hemorrhage. A. Actively bleeding sigmoid diverticulum. B. Treatment of the bleeding vessel at the dome of the diverticulum with a contact thermal probe. C. Hemostasis secured following contact coagulation with tattoo injection to aid future localization.



FIGURE 322-44 Esophageal food impaction. Meat bolus impacted in the distal esophagus.

DYSPHAGIA

About 50% of patients presenting with difficulty swallowing have a mechanical obstruction; the remainder has a motility disorder, such as achalasia or diffuse esophageal spasm. Careful history-taking often points to a presumptive diagnosis and leads to the appropriate use of diagnostic tests. Esophageal strictures (Fig. 322-53) typically cause progressive dysphagia, first for solids, then for liquids; motility disorders often cause intermittent dysphagia for both solids and liquids. Some underlying disorders have characteristic historic features: Schatzki's ring (Fig. 322-54) causes episodic dysphagia for solids, typically at the beginning of a meal; oropharyngeal motor disorders typically present with difficulty initiating deglutition (*transfer dysphagia*) and nasal reflux or coughing with swallowing; and achalasia may cause nocturnal regurgitation of undigested food.

When mechanical obstruction is suspected, endoscopy is a useful initial diagnostic test, since it permits immediate biopsy and/or dilation of strictures, masses, or rings. The presence of linear furrows and multiple corrugated rings throughout a narrowed esophagus should raise suspicion for eosinophilic esophagitis, an increasingly recognized cause of recurrent dysphagia and food impaction (Fig. 322-55). Blind



FIGURE 322-45 Esophageal foreign body. Intentionally ingested toothbrush impacted in the esophageal lumen.



A



B



C

FIGURE 322-46 Gastric outlet obstruction due to pyloric stenosis. A. Nonsteroidal anti-inflammatory agent-induced ulcer disease with severe stenosis of the pylorus (arrow). B. Balloon dilation of the stenosis. C. Appearance of pyloric ring after dilation.

or forceful passage of an endoscope may lead to perforation in a patient with stenosis of the cervical esophagus or a Zenker's diverticulum (Fig. 322-56), but gentle passage of an endoscope under direct visual guidance is reasonably safe. Endoscopy can miss a subtle stricture or ring in some patients.

When transfer dysphagia is evident or an esophageal motility disorder is suspected, esophageal radiography and/or a video-swallow study are the best initial diagnostic tests. The oropharyngeal swallowing mechanism, esophageal peristalsis, and the lower esophageal sphincter can all be assessed. In some disorders, subsequent esophageal manometry is required for diagnosis.

Various causes of dysphagia are amenable to endoscopic therapy. Benign strictures, rings, and webs can be dilated using a



FIGURE 322-47 Sigmoid volvulus with the characteristic radiologic appearance of a “bent inner tube.”

through-the-scope balloon (Fig. 322-57) or a polyvinyl dilator passed over a guide wire. In some instances, fibrotic strictures may respond to needle-knife electroincision (Fig. 322-58) when they prove refractory to dilation. Self-expanding esophageal stents can be used to palliate dysphagia from malignant obstruction (Fig. 322-59), and flexible endoscopic myotomy is an option for Zenker’s diverticulum (Video V5-19). Recent advances in submucosal endoscopy have enabled the development of procedures, such as POEM (Video V5-20) and POET (Video V5-21), for the management of achalasia and select subepithelial esophageal tumors, respectively.

■ ENDOSCOPIC TREATMENT OF OBESITY

A significant proportion of Americans are overweight or obese, and obesity-associated diabetes has become a major public health



A



B

FIGURE 322-48 Acute colonic pseudoobstruction. A. Acute colonic dilation occurring in a patient soon after knee surgery. B. Colonoscopic placement of decompression tube with marked improvement in colonic dilation.



A



B



C

FIGURE 322-49 Obstructing colonic carcinoma. A. Colonic adenocarcinoma causing marked luminal narrowing of the distal transverse colon. B. Endoscopic placement of a self-expandable metal stent. C. Radiograph of expanded stent across the obstructing tumor with a residual waist (arrow).

problem. Bariatric surgery is the most effective weight-loss intervention, decreasing long-term mortality in obese persons, but many patients do not undergo surgery. Endoscopic treatments for obesity have been developed and include insertion of an intragastric balloon or duodenojejunral bypass liner, placement of a percutaneous gastric tube for aspiration of gastric contents after meals, or endoscopic sleeve gastroplasty, which utilizes endoscopic suturing to narrow the lumen of the gastric body (Video V5-22). Prospective trials show that these treatments induce total-body weight loss of 7–20% and provide varying degrees of glycemic control. Additional endoscopic modalities are

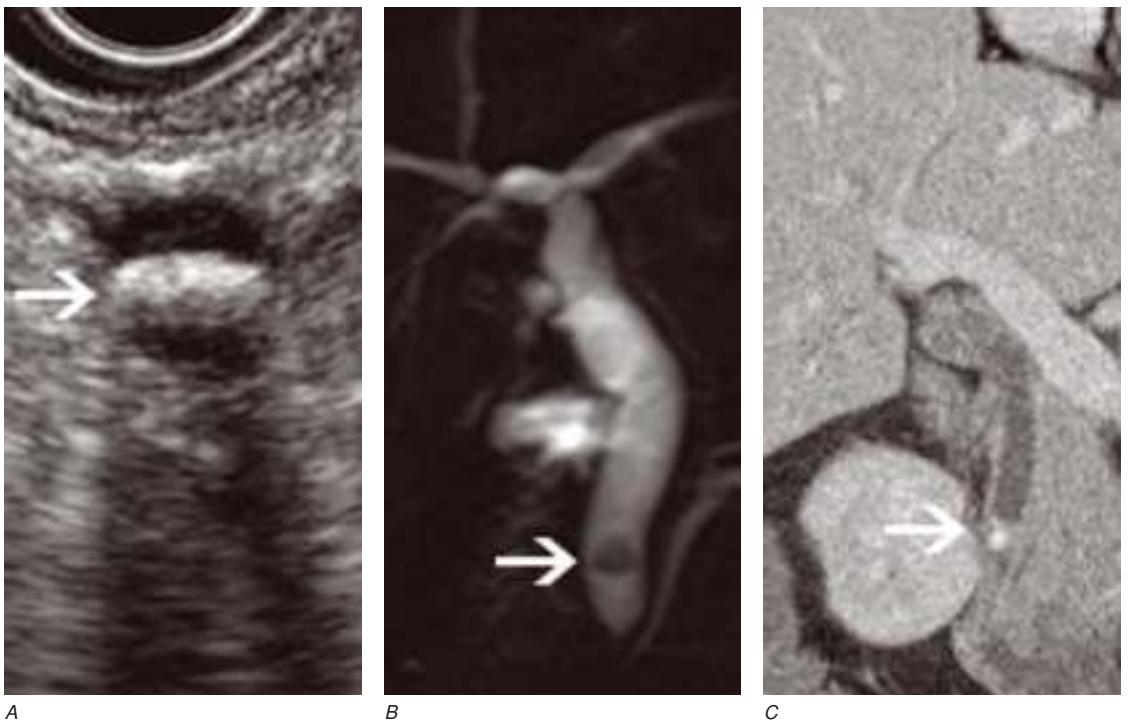


FIGURE 322-50 Methods of bile duct imaging. Arrows mark bile duct stones. A. Endoscopic ultrasound (EUS). B. Magnetic resonance cholangiopancreatography (MRCP). C. Helical computed tomography (CT).

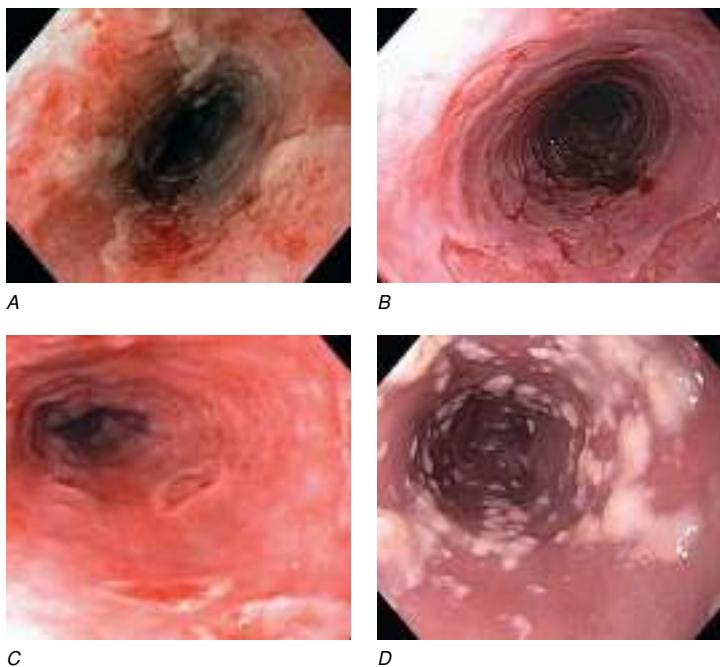


FIGURE 322-51 Causes of esophagitis. A. Severe reflux esophagitis with mucosal ulceration and friability. B. Cytomegalovirus esophagitis. C. Herpes simplex virus esophagitis with target-type shallow ulcerations. D. *Candida* esophagitis with white plaques adherent to the esophageal mucosa.

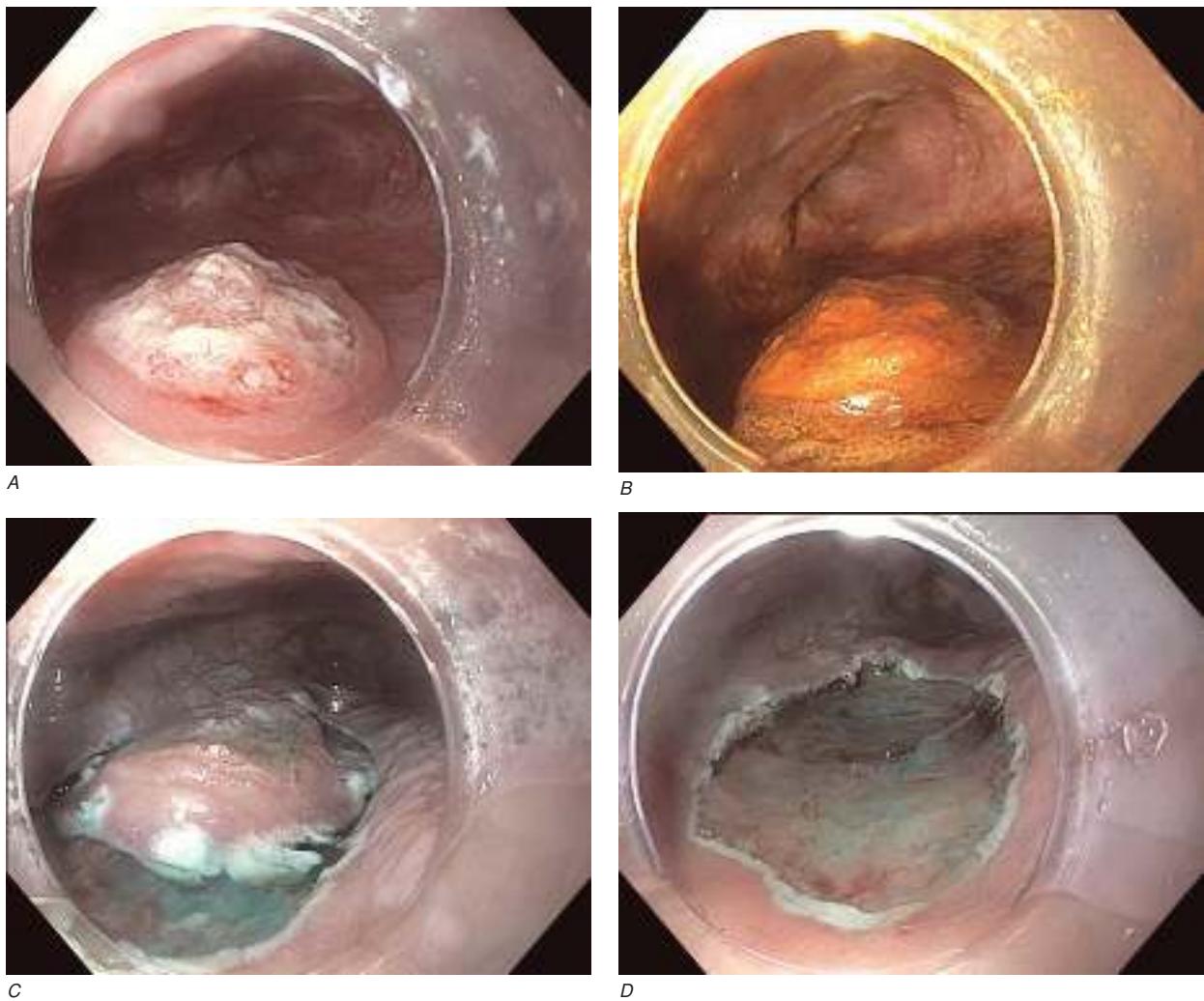


FIGURE 322-52 Early squamous cell cancer. *A*, Nodularity in the distal esophagus due to T1 esophageal squamous cell cancer. *B*, Lesion is unstained under Lugol's iodine chromoendoscopy without additional unstained areas. *C*, Circumferential mucosal incision around the lesion. *D*, Resection defect following en bloc removal of the lesion via endoscopic submucosal dissection.

undergoing clinical trials. The long-term efficacy of endoscopic bariatric treatment in comparison to surgery is still unclear.

TREATMENT OF MALIGNANCIES

Endoscopy plays an important role in the treatment of gastrointestinal malignancies. Early-stage malignancies limited to the mucosal and superficial submucosal layers may be resected using the techniques of EMR (Video V5-4) or ESD (Video V5-5). RFA and cryotherapy are effective modalities for ablative treatment of high-grade dysplasia and intramucosal cancer in Barrett's esophagus (Video V5-23). Gastrointestinal stromal tumors can be removed en bloc by EFTR (Video V5-3). In general, endoscopic techniques offer the advantage of a minimally invasive approach to treatment but rely on other imaging techniques (such as CT, MRI, positron emission tomography [PET], and EUS) to exclude distant metastases or locally advanced disease better treated by surgery or other modalities. The decision to treat an early-stage gastrointestinal malignancy endoscopically is often made in collaboration with a surgeon and/or oncologist.

Endoscopic palliation of gastrointestinal malignancies relieves symptoms and, in many cases, prolongs survival. Malignant obstruction can be relieved by endoscopic stent placement (Figs. 322-18, 322-49,

322-59, and 322-60; Videos V5-16 and V5-17), and malignant gastrointestinal bleeding can often be palliated endoscopically as well. EUS-guided celiac plexus neurolysis may relieve pancreatic cancer pain.

ANEMIA AND OCCULT BLOOD IN THE STOOL

Iron-deficiency anemia may be attributed to poor iron absorption (as in celiac sprue) or, more commonly, chronic blood loss. Intestinal bleeding should be strongly suspected in men and postmenopausal women with iron-deficiency anemia, and colonoscopy is indicated in such patients, even in the absence of detectable occult blood in the stool. Approximately 30% will have large colonic polyps or colorectal cancer, and a few patients will have colonic vascular lesions. When a convincing source of blood loss is not found in the colon, upper gastrointestinal endoscopy should be considered; if no lesion is found, duodenal biopsies should be obtained to exclude sprue (Fig. 322-61). Small-bowel evaluation with capsule endoscopy (Fig. 322-62), CT or magnetic resonance (MR) enterography, or device-assisted enteroscopy may be appropriate if both EGD and colonoscopy are unrevealing.

Tests for occult blood in the stool detect hemoglobin or the heme moiety and are most sensitive for colonic blood loss, although they will also detect larger amounts of upper gastrointestinal bleeding. Patients



FIGURE 322-53 Peptic esophageal stricture associated with esophagitis.

with occult blood in the stool should undergo colonoscopy to diagnose or exclude colorectal neoplasia, especially if they are >50 years old or have a family history of colonic neoplasia. Whether upper endoscopy is also indicated depends on the patient's symptoms.

The small intestine may be the source of chronic intestinal bleeding, especially if colonoscopy and upper endoscopy are not diagnostic. The utility of small-bowel evaluation varies with the clinical setting and is most important in patients in whom bleeding causes chronic or recurrent anemia. In contrast to the low diagnostic yield of small-bowel radiography, positive findings on capsule endoscopy are seen in 50–70% of patients with suspected small-intestinal bleeding. The most common finding is mucosal vascular ectasia. CT and MR enterography accurately detect small-bowel masses and Crohn's disease and are also useful for initial small-bowel evaluation. Deep enteroscopy may follow capsule endoscopy for biopsy of lesions or to provide specific therapy, such as argon plasma coagulation of vascular ectasias (Fig. 322-63).

■ COLORECTAL CANCER SCREENING

The majority of colon cancers develop from preexisting colonic adenomas, and colorectal cancer can be largely prevented by the detection and removal of adenomatous polyps (Video V5-24). The choice of screening strategy for an asymptomatic person depends on personal



FIGURE 322-55 Eosinophilic esophagitis. Multiple circular rings of the esophagus creating a corrugated appearance and an impacted grape at the narrowed esophagogastric junction. The diagnosis requires biopsy with histologic finding of >15–20 eosinophils/high-power field.

and family history. Individuals with inflammatory bowel disease, a history of colorectal polyps or cancer, family members with adenomatous polyps or cancer, or certain familial cancer syndromes (Fig. 322-64) are at increased risk for colorectal cancer. An individual without these factors is generally considered at average risk.

Screening strategies are summarized in Table 322-3. While fecal immunochemical tests (FIT) for heme or stool tests for occult blood have been shown to decrease the mortality rate from colorectal cancer, they do not detect some cancers and many polyps. FIT-DNA multitar geted stool DNA tests appear to be more sensitive, but direct visualization of the colon is the gold standard method for detection of polyps and cancers and remains a preferred screening strategy. Sigmoidoscopy is also used for colorectal cancer screening. However, the distribution of colon cancers has changed in the United States over time, with proportionally fewer rectal and left-sided cancers than in the past. Large American studies of colonoscopy for screening of average-risk individuals show that cancers are roughly equally distributed between the left and right colon and half of patients with right-sided lesions have no polyps in the left colon. Visualization of the entire colon thus appears to be the optimal strategy for colorectal cancer screening and prevention.

Computed tomography colonography (CTC) is a radiologic technique that images the colon with CT following rectal insufflation of the colonic lumen. Computer rendering of CT images generates an electronic display of a virtual “flight” along the colonic lumen, simulating colonoscopy (Fig. 322-65). Findings detected during CTC often require subsequent conventional colonoscopy for confirmation and treatment.

■ DIARRHEA

Most cases of diarrhea are acute, self-limited, and due to infections or medication. Chronic diarrhea (lasting >6 weeks) is more often due to a primary inflammatory, malabsorptive, or motility disorder; is less likely to resolve spontaneously; and generally requires diagnostic evaluation. Patients with chronic diarrhea or severe, unexplained acute diarrhea often undergo endoscopy if stool tests for pathogens are unrevealing. The choice of endoscopic testing depends on the clinical setting.

Patients with colonic symptoms and findings such as bloody diarrhea, tenesmus, fever, or leukocytes in stool generally undergo sigmoidoscopy or colonoscopy to assess for colitis (Fig. 322-9). Sigmoidoscopy is an appropriate initial test in most patients. Conversely, patients with symptoms and findings suggesting small-bowel disease,



FIGURE 322-54 Schatzki's ring at the gastroesophageal junction.

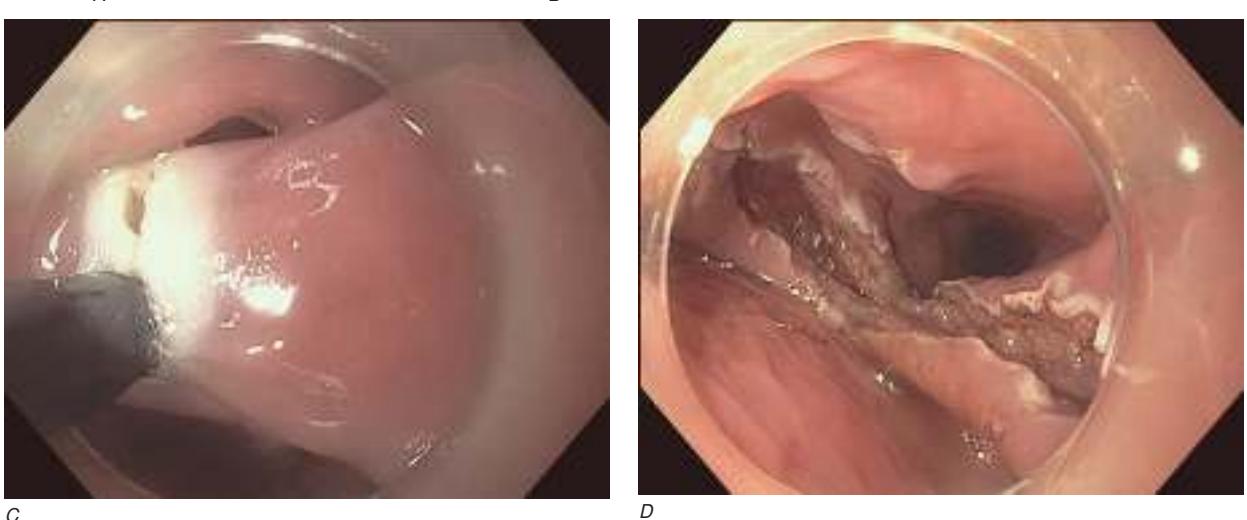
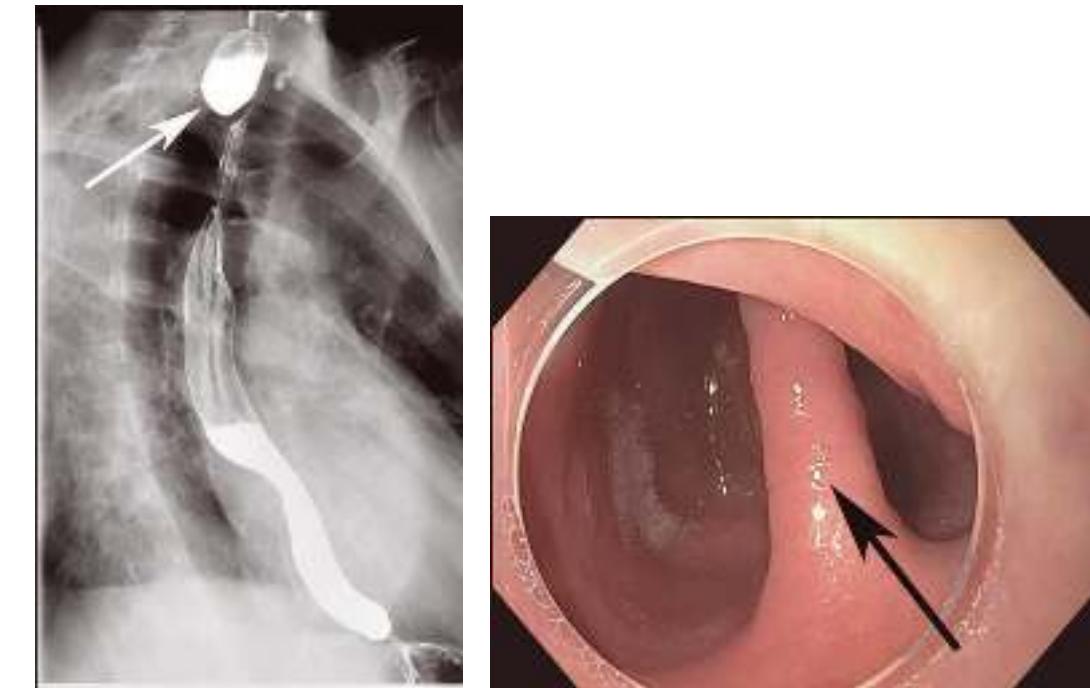


FIGURE 322-56 Zenker's diverticulum. *A*, Contrast esophagography demonstrates a moderate-sized Zenker's diverticulum. *B*, Endoscopic view of the Zenker's diverticulum (left) relative to the true esophageal lumen (right) separated by the diverticular septum. *C*, Flexible endoscopic diverticulotomy using an electrosurgical knife. *D*, Appearance post diverticulotomy.



FIGURE 322-57 Endoscopic management of peptic stricture. *A*, Peptic stricture. *B*, Through-the-scope balloon dilation of stricture. *C*, Improvement in luminal diameter after dilation.



FIGURE 322-58 Endoscopic management of an esophagogastric anastomotic stricture. *A*, Recurrent anastomotic stricture despite periodic balloon dilation. *B*, Needle-knife electroincision of stricture. *C*, Improvement in luminal opening after therapy.

such as large-volume watery stools, substantial weight loss, and malabsorption of iron, calcium, or fat, may undergo upper endoscopy with duodenal aspirates for assessment of bacterial overgrowth and biopsies for assessment of mucosal diseases, such as celiac sprue.

Many patients with chronic diarrhea do not fit either of these patterns. In the setting of a long-standing history of alternating constipation and diarrhea dating to early adulthood, without findings such as blood in the stool or anemia, a diagnosis of irritable bowel syndrome may be made without direct visualization of the bowel. Steatorrhea and upper abdominal pain may prompt evaluation of the pancreas rather than the gut. Patients whose chronic diarrhea is not easily categorized often undergo initial colonoscopy to examine the entire colon and terminal ileum for inflammatory or neoplastic disease (Fig. 322-66).

■ MINOR HEMATOchezIA

Bright red blood passed with or on formed brown stool usually has an anal, rectal, or sigmoid source (Fig. 322-67). Even trivial amounts of hematochezia should be investigated with colonoscopy and/or flexible sigmoidoscopy together with anoscopy to exclude polyps or cancers, especially in patients >40 years old and those with a personal or family history of colorectal polyps or cancer. Patients reporting red blood on the toilet tissue only, without blood in the toilet or on the stool, are generally bleeding from a lesion in the anal canal; careful external inspection, digital examination, and sigmoidoscopy with anoscopy may be sufficient for diagnosis in such cases.

■ PANCREATITIS

About 20% of patients with pancreatitis have no identified cause after routine clinical investigation (including a review of medication and alcohol use; measurement of serum triglyceride, calcium, and

immunoglobulin G subclass 4 levels; abdominal ultrasonography; and CT or MRI). Endoscopic assessment leads to a specific diagnosis in the majority of such patients, often altering clinical management. Endoscopic investigation is particularly appropriate if the patient has had more than one episode of pancreatitis.

Microlithiasis, or the presence of microscopic crystals in bile, is a leading cause of previously unexplained acute pancreatitis and is sometimes seen during abdominal ultrasonography as layering sludge or flecks of floating, echogenic material in the gallbladder. EUS may identify previously undetected microlithiasis.

Previously undetected chronic pancreatitis, pancreatic malignancy, or pancreas divisum may be diagnosed by either ERCP or EUS. Autoimmune pancreatitis is often suspected based on CT, MRI, or serologic findings, but it may first become apparent during EUS and may require EUS-guided pancreatic biopsy for histologic diagnosis.

Severe pancreatitis often results in pancreatic fluid collections. Symptomatic pseudocysts and areas of walled-off pancreatic necrosis can be drained into the stomach or duodenum endoscopically, using transpapillary and transmural endoscopic techniques. Pancreatic necrosis can be debrided by direct endoscopic necrosectomy (Video V5-2) via an endoscopically created transmural drainage site.

■ CANCER STAGING

Local staging of esophageal, gastric, pancreatic, bile duct, and rectal cancers can be obtained with EUS (Fig. 322-20). EUS with fine-needle aspiration (Fig. 322-21) currently provides the most accurate preoperative assessment of local tumor and nodal staging, but it does not detect many distant metastases. Details of the local tumor stage can guide treatment decisions including resectability and need for neoadjuvant

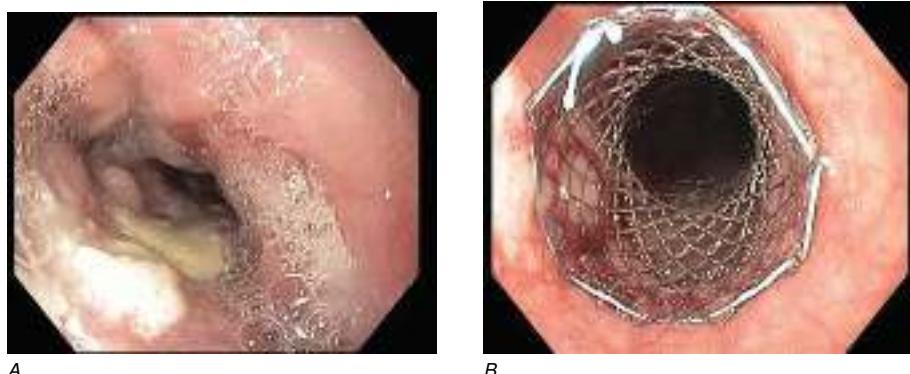


FIGURE 322-59 Palliation of malignant dysphagia. *A*, Obstructing distal esophageal cancer. *B*, Palliative stent placement.

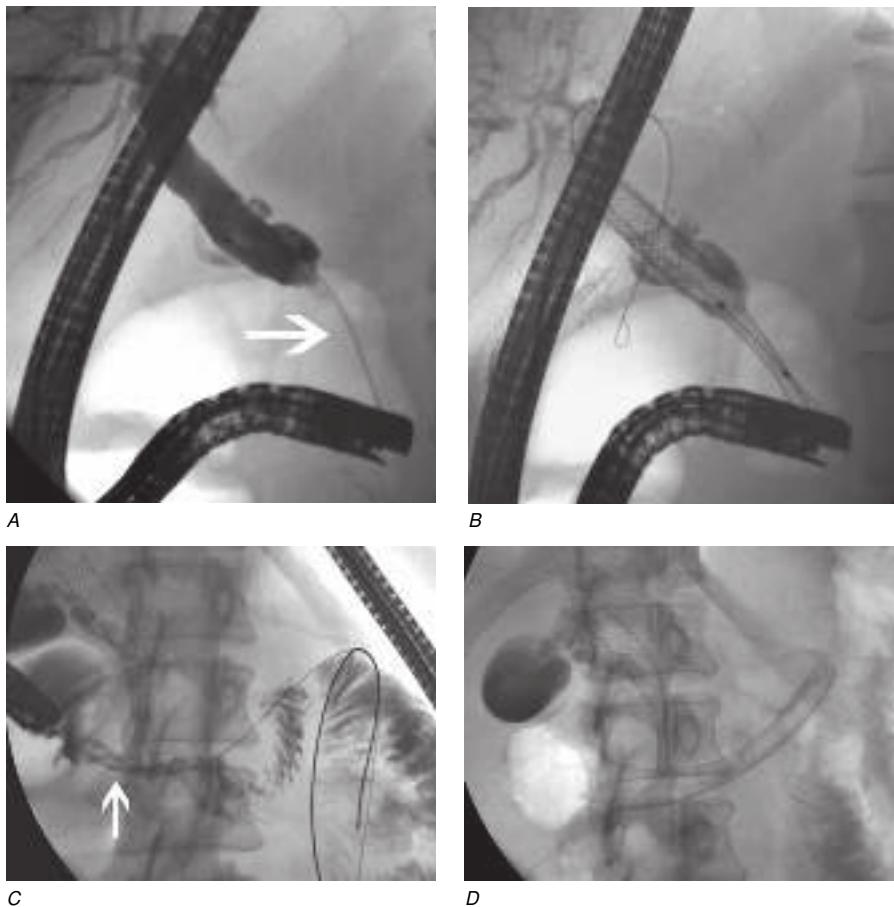


FIGURE 322-60 Placement of biliary and duodenal self-expanding metal stents (SEMS) for obstruction caused by pancreatic cancer. *A*, Endoscopic retrograde cholangiopancreatography (ERCP) demonstrates a distal bile duct stricture (arrow). *B*, A biliary SEMS is placed. *C*, Contrast injection demonstrates a duodenal stricture (arrow). *D*, Biliary and duodenal SEMS in place.



FIGURE 322-61 Celiac sprue. Scalloped duodenal folds in a patient with celiac sprue.

therapy. EUS with transesophageal needle biopsy may also be used to assess the presence of non-small-cell lung cancer in mediastinal nodes.

OPEN ACCESS ENDOSCOPY

Direct scheduling of endoscopic procedures by primary care physicians without preceding gastroenterology consultation, or *open-access endoscopy*, is common. When the indications for endoscopy are clear-cut and appropriate, the procedural risks are low, and the patient understands what to expect, open-access endoscopy streamlines patient care and decreases costs.

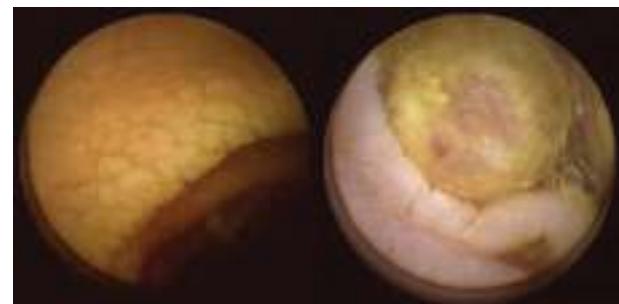


FIGURE 322-62 Capsule endoscopy. Images of a mildly scalloped jejunal fold (left) and an ileal tumor (right) in a patient with celiac sprue. (Images courtesy of Dr. Elizabeth Rajan; with permission.)

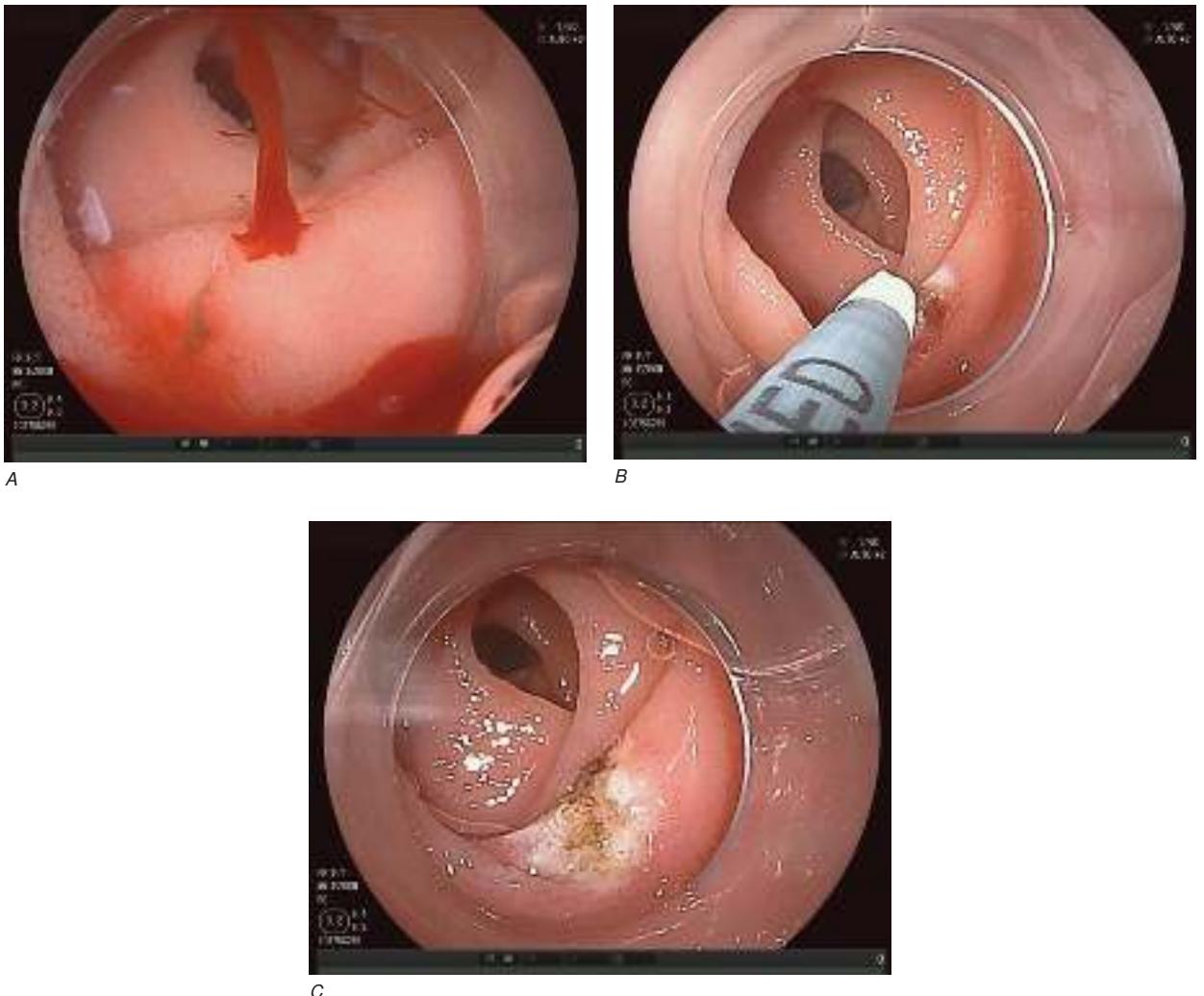


FIGURE 322-63 Small-bowel vascular ectasia. *A*. Actively bleeding mid-jejunal vascular ectasia identified by double-balloon enteroscopy. *B*. Ablation of vascular ectasia with argon plasma coagulation (APC). *C*. Hemostasis secured following APC.



FIGURE 322-64 Familial adenomatous polyposis. Numerous colon polyps in a patient with familial adenomatous polyposis syndrome.

Patients referred for open-access endoscopy should have a recent history, physical examination, and medication list that are available for review when the patient comes to the endoscopy suite. Patients with unstable or symptomatic cardiovascular or respiratory conditions should not be referred directly for open-access endoscopy. Those with particular conditions who are undergoing certain procedures should be prescribed prophylactic antibiotics prior to endoscopy (Table 322-1). In addition, patients taking anticoagulants and/or antiplatelet drugs may require adjustment of these agents before endoscopy based on the procedural risk for bleeding and their underlying risk for a thromboembolic event (Table 322-2).

Common indications for open-access EGD include dyspepsia resistant to a trial of appropriate therapy, dysphagia, gastrointestinal bleeding, and persistent anorexia or early satiety. Open-access colonoscopy is often requested in men or postmenopausal women with iron-deficiency anemia, in patients with hematochezia or occult blood in the stool, in patients with a previous history of colorectal adenomatous polyps or cancer, and for colorectal cancer screening. Flexible sigmoidoscopy is commonly performed as an open-access procedure.

TABLE 322-3 Colorectal Cancer Screening Strategies

	CHOICES/RECOMMENDATIONS	COMMENTS
Average-Risk Patients		
Asymptomatic individuals ≥45 years old	Colonoscopy every 10 years ^a Multitargeted stool DNA test every 3 years Annual FIT or FOBT, multiple take-home specimen cards, with or without sigmoidoscopy every 5–10 years CT colonography every 5 years Flexible sigmoidoscopy every 5 years	Preferred cancer prevention strategy Less sensitive than colonoscopy; colonoscopy if results are positive Does not detect many polyps; colonoscopy if results are positive Colonoscopy if results are positive Does not detect proximal colon polyps and cancers; colonoscopy if an adenomatous polyp is found
Personal History of Polyps or CRC		
1–2 small (<1 cm) adenomas with low-grade dysplasia	Repeat colonoscopy in 5–10 years ^a	Assuming complete polyp resection. Interval may vary based on prior personal history and family history
3–10 adenomas, or any high-risk adenoma ^b	Repeat colonoscopy in 3 years ^a ; subsequent colonoscopy based on findings	Assuming complete polyp resection
>10 adenomas	Repeat colonoscopy in <3 years based on clinical judgment ^a	Consider evaluation for FAP or HNPCC; see recommendations below
Piecemeal removal of a sessile polyp	Exam in 2–6 months to verify complete removal	
Small (<1 cm) hyperplastic polyps of sigmoid and rectum	Repeat colonoscopy in 10 years ^a	Those with hyperplastic polyposis syndrome merit more frequent follow-up
Sessile serrated adenoma/polyp <10 mm, without dysplasia	Repeat colonoscopy in 5 years ^a	
Sessile serrated adenoma/polyp ≥10 mm or with dysplasia, or ≥2 serrated polyps	Repeat colonoscopy in 3 years ^a	Serrated polyposis syndrome merits more frequent follow-up
Incompletely removed serrated polyp ≥1 cm	Exam in 2–6 months to verify complete removal	
Colon cancer	Evaluate entire colon around the time of resection, then repeat colonoscopy in 1 year ^a	Subsequent colonoscopy in 3 years if the 1-year examination is normal
Inflammatory Bowel Disease		
Long-standing (>8 years) ulcerative pancolitis or Crohn's colitis, or left-sided ulcerative colitis of >15 years' duration	Colonoscopy with biopsies every 1–2 years	Consider chromoendoscopy or other advanced imaging techniques for detection of flat dysplasia during colonoscopy
Family History of Polyps or CRC		
First-degree relatives with only small tubular adenomas	Same as average risk	
One first-degree relative with CRC or advanced adenoma at age ≥60 years	Colonoscopy every 10 years starting at age 40	
One first-degree relative with CRC or advanced adenoma at age <60 years, or two first-degree relatives with CRC or advanced adenomas at any age	Colonoscopy every 5 years beginning at age 40 years or 10 years younger than age at diagnosis of the youngest affected relative, whichever is earlier	
Familial adenomatous polyposis (FAP)	Sigmoidoscopy or colonoscopy annually, beginning at age 10–12 years	Consider genetic counseling and testing; consider screening family members
Hereditary nonpolyposis colorectal cancer (HNPCC; Lynch syndrome)	Colonoscopy every 2 years beginning at age 20–25 years (or 10 years younger than the youngest first-degree relative was when diagnosed with CRC) until age 40, then annually thereafter	Consider histologic evaluation for microsatellite instability in tumor specimens of patients who meet modified Bethesda criteria; consider genetic counseling and testing, consider screening family members
Serrated polyposis syndrome (SPS)	Colonoscopy at age 40 (or the same age at which the youngest first-degree relative was when diagnosed with SPS, or 10 years younger than the youngest first-degree relative was when diagnosed with CRC), then every 1–2 years thereafter	Consider screening family members, even of patients with multiple serrated polyps who do not meet SPS criteria.

^a Assumes good colonic preparation and complete examination to cecum. ^b High-risk adenoma: any adenoma ≥1 cm in size or containing high-grade dysplasia or villous features.

Abbreviations: CRC, colorectal cancer; FIT, fecal immunochemical test; FOBT, fecal occult blood test.

Sources: Adapted from U.S. Preventative Services Task Force Draft Guidelines released in 2020 (<https://uspreventiveservicestaskforce.org/uspstf/draft-recommendation/colorectal-cancer-screening>) and American Cancer Society Guidelines (<https://www.cancer.org/cancer/colon-rectal-cancer/detection-diagnosis-staging/acs-recommendations.html>), both accessed on December 12, 2020. See also G Mankani et al: Serrated polyposis syndrome. *Clin Gastroenterol Hepatol* 18:777, 2020.

When patients are referred for open-access colonoscopy, the primary care provider may need to choose a colonic preparation. Commonly used oral preparations include polyethylene glycol lavage solution, with or without citric acid. A “split-dose” regimen improves

the quality of colonic preparation. Osmotic purgative preparations (such as sodium phosphate) are also effective but may cause fluid and electrolyte abnormalities and renal toxicity, especially in patients with renal failure or congestive heart failure and those >70 years of age.



FIGURE 322-65 Virtual colonoscopy image of a colon polyp (arrow). (Image courtesy of Dr. Jeff Fidler; with permission.)

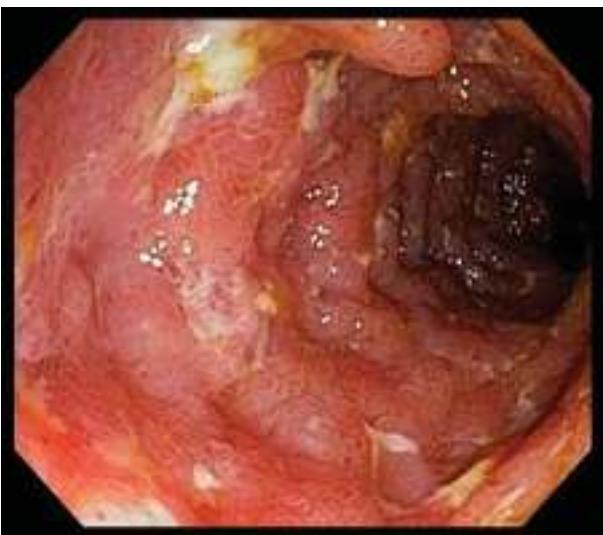


FIGURE 322-66 Crohn's ileitis. Edema, erythema, ulcers, and exudates involving the terminal ileum.



FIGURE 322-67 Internal hemorrhoids with bleeding stigmata (arrow) as seen on retroflexed view of the rectum.

FURTHER READING

- ASGE S P C et al: Antibiotic prophylaxis for GI endoscopy. *Gastrointest Endosc* 81:81, 2015.
- ASGE S P C et al: Open-access endoscopy. *Gastrointest Endosc* 81:1326, 2015.
- B AN et al: Management of nonvariceal upper gastrointestinal bleeding: Guideline recommendations from the international consensus group. *Ann Intern Med* 171:805, 2019.
- G -T 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.
- R DK et al: Colorectal cancer screening: Recommendations for physicians and patients from the U.S. Multi-Society Task Force on Colorectal Cancer. *Gastroenterology* 153:307, 2017.
- S NJ et al: ACG clinical guideline: Diagnosis and management of Barrett's esophagus. *Am J Gastroenterol* 111:30, 2016.
- S LL et al: ACG clinical guideline: Management of patients with acute lower gastrointestinal bleeding. *Am J Gastroenterol* 111:459, 2016.

323

Diseases of the Esophagus

Peter J. Kahrilas, Ikuo Hirano



ESOPHAGEAL STRUCTURE AND FUNCTION

The esophagus is a hollow, muscular tube coursing through the posterior mediastinum joining the hypopharynx to the stomach with a sphincter at each end. It functions to transport food and fluid between these ends, otherwise remaining empty. The physiology of swallowing, esophageal motility, and oral and pharyngeal dysphagia are described in Chap. 44. Esophageal diseases can be manifested by impaired function or pain. Key functional impairments are swallowing disorders and excessive gastroesophageal reflux. Pain, sometimes indistinguishable from cardiac chest pain, can result from inflammation, infection, dysmotility, or neoplasm.

SYMPTOMS OF ESOPHAGEAL DISEASE

The clinical history remains central to the evaluation of esophageal symptoms. A thoughtfully obtained history will often expedite management. Important details include weight gain or loss, gastrointestinal bleeding, dietary habits including the timing of meals, smoking, and alcohol consumption. The major esophageal symptoms are heartburn, regurgitation, chest pain, dysphagia, odynophagia, and globus sensation.

Heartburn (pyrosis), the most common esophageal symptom, is characterized by a discomfort or burning sensation behind the sternum that arises from the epigastrium and may radiate toward the neck. Heartburn is an intermittent symptom, most commonly experienced after eating, during exercise, and while lying recumbent. The discomfort is relieved with drinking water or taking an antacid but can occur frequently, interfering with normal activities including sleep. The association between heartburn and gastroesophageal reflux disease (GERD) is so strong that empirical therapy for GERD has become accepted management. However, the term *heartburn* is often misused and/or referred to using other terms such as *indigestion* or *repeating*, making it important to clarify the intended meaning.

Regurgitation is the effortless return of food or fluid into the pharynx without nausea or retching. Patients report a sour or burning fluid in the throat or mouth that may also contain undigested food particles. Bending, belching, or maneuvers that increase intraabdominal

pressure can provoke regurgitation. A clinician needs to discriminate among regurgitation, vomiting, and rumination. *Vomiting* is preceded by nausea and accompanied by retching. *Rumination* is a behavior in which recently swallowed food is regurgitated and then reswallowed repetitively for up to an hour. Although there is some linkage between rumination and cognitive deficiency, the behavior is also exhibited by unimpaired individuals.

Chest pain is a common esophageal symptom with characteristics similar to cardiac pain, sometimes making this distinction difficult. Esophageal pain is usually experienced as a pressure-type sensation in the mid chest, radiating to the mid back, arms, or jaws. The similarity to cardiac pain is likely because the two organs share a nerve plexus and the nerve endings in the esophageal wall have poor discriminative ability among stimuli. Esophageal distention or even chemostimulation (e.g., with acid) will often be perceived as chest pain. Gastroesophageal reflux is the most common cause of esophageal chest pain.

Esophageal *dysphagia* (Chap. 44) is often described as a feeling of food “sticking” or even lodging in the chest. Important distinctions are between uniquely solid food dysphagia as opposed to liquid and solid, episodic versus constant dysphagia, and progressive versus static dysphagia. If the dysphagia is for liquids as well as solid food, it suggests a motility disorder such as achalasia. Conversely, uniquely solid food dysphagia is suggestive of a stricture, ring, or tumor. Of note, a patient’s localization of food hang-up in the esophagus is notoriously imprecise. Approximately 30% of distal esophageal obstructions are perceived as cervical dysphagia. In such instances, the absence of concomitant symptoms generally associated with oropharyngeal dysphagia such as aspiration, nasopharyngeal regurgitation, cough, drooling, or obvious neuromuscular compromise should suggest an esophageal etiology.

Odynophagia is pain either caused by or exacerbated by swallowing. Although typically considered distinct from dysphagia, odynophagia may manifest concurrently with dysphagia. Odynophagia is more common with pill or infectious esophagitis than with reflux esophagitis and should prompt a search for these entities. When odynophagia does occur in GERD, it is likely related to an esophageal ulcer or extensive erosions.

Globus sensation, also known as globus pharyngeus, is the perception of a lump or fullness in the throat that is felt irrespective of swallowing. Although such patients are frequently referred for an evaluation of dysphagia, globus sensation is often relieved by the act of swallowing. As implied by its alternative name, “globus hystericus,” globus sensation often occurs in the setting of anxiety or obsessive-compulsive disorders. Clinical experience teaches that it is often attributable to GERD.

Water brash is excessive salivation resulting from a vagal reflex triggered by acidification of the esophageal mucosa. This is not a common symptom. Afflicted individuals will describe the unpleasant sensation of the mouth rapidly filling with salty thin fluid, often in the setting of concomitant heartburn.

DIAGNOSTIC STUDIES

■ ENDOSCOPY

Endoscopy, also known as esophagogastroduodenoscopy (EGD), is the most useful test for the evaluation of the proximal gastrointestinal tract. Modern instruments produce high-quality, color images of the esophageal, gastric, and duodenal lumen. Endoscopes also have an instrumentation channel through which biopsy forceps, injection catheters for local delivery of therapeutic agents, balloon dilators, or devices for hemostasis or removal of mucosal lesions can be used. The key advantages of endoscopy over barium radiography are as follows: (1) increased sensitivity for the detection of mucosal lesions; (2) vastly increased sensitivity for the detection of abnormalities mainly identifiable by color, such as Barrett’s metaplasia or vascular lesions; (3) the ability to obtain biopsy specimens for histologic examination of suspected abnormalities; and (4) the ability to dilate strictures during the examination. Submucosal endoscopy has emerged as a diagnostic modality for assessment of subepithelial lesions and therapy of esophageal motility disorders. The main disadvantages of endoscopy are low

sensitivity for detection of diffuse, nonfocal esophageal strictures, cost, and the need for sedatives or anesthetics.

■ RADIOGRAPHY

Contrast radiography of the esophagus, stomach, and duodenum can demonstrate reflux of the contrast media, hiatal hernia, mucosal granularity, erosions, ulcerations, and strictures. The sensitivity of radiography compared with endoscopy for detecting reflux esophagitis reportedly ranges from 22 to 95%, with higher grades of esophagitis (i.e., ulceration or stricture) exhibiting greater detection rates. Conversely, the sensitivity of barium radiography for detecting esophageal strictures is greater than that of endoscopy, especially when the study is done in conjunction with a 13-mm barium tablet. Barium studies also provide an assessment of esophageal function and morphology that may be undetected on endoscopy. Tracheoesophageal fistula, altered postsurgical anatomy, and extrinsic esophageal compression are conditions where radiographic imaging complements endoscopic assessment. Hypopharyngeal pathology and disorders of the cricopharyngeus muscle are better appreciated on radiographic examination than with endoscopy, particularly with rapid sequence or video fluoroscopic recording. The major shortcoming of barium radiography is that it rarely obviates the need for endoscopy. Either a positive or a negative study is usually followed by an endoscopic evaluation to obtain biopsies, provide therapy, or clarify findings in the case of a positive examination or to add a level of certainty in the case of a negative examination.

■ ENDOSCOPIC ULTRASOUND

Endoscopic ultrasound (EUS) instruments combine an endoscope with an ultrasound transducer to create a transmural image of the tissue surrounding the endoscope tip. The key advantage of EUS over alternative radiologic imaging techniques is much greater resolution attributable to the proximity of the ultrasound transducer to the area being examined. Available devices can provide either radial imaging (360-degree, cross-sectional) or a curved linear image that can guide fine-needle aspiration of imaged structures such as lymph nodes or tumors. Major esophageal applications of EUS are to stage esophageal cancer, to evaluate dysplasia in Barrett’s esophagus, and to assess submucosal lesions.

■ ESOPHAGEAL MANOMETRY

Esophageal manometry, or motility testing, entails positioning a pressure-sensing catheter within the esophagus and then observing the contractility following test swallows. The upper esophageal sphincter and lower esophageal sphincter (LES) appear as zones of high pressure that relax on swallowing, whereas the intersphincteric esophagus exhibits peristaltic contractions. Manometry is used to diagnose motility disorders (achalasia, diffuse esophageal spasm [DES]) and to assess peristaltic integrity prior to the surgery for reflux disease. Technologic advances have enhanced esophageal manometry as high-resolution esophageal pressure topography (Fig. 323-1). Manometry can also be combined with intraluminal impedance monitoring. Impedance recordings use a series of paired electrodes added to the manometry catheter. Esophageal luminal contents in contact with the electrodes decrease (liquid) or increase (air) the impedance signal, allowing detection of anterograde or retrograde esophageal bolus transit.

■ REFLUX TESTING

GERD is often diagnosed in the absence of endoscopic signs of esophagitis, which would otherwise define the disease. This occurs in the settings of partially treated disease, an abnormally sensitive esophageal mucosa, or, most commonly, in nonerosive reflux disease. In such instances, reflux testing can demonstrate excessive esophageal exposure to refluxed gastric fluid, the physiologic abnormality of GERD. This can be done by ambulatory 24- to 96-h esophageal pH recording using either a wireless pH-sensitive transmitter that is affixed to the esophageal mucosa or a transnasally positioned wire electrode with the tip stationed in the distal esophagus. Either way, the outcome is expressed as the percentage of the day that the pH was <4 (indicative of recent acid reflux), with values exceeding 5% indicative of GERD. Reflux testing is useful in the evaluation of patients presenting with

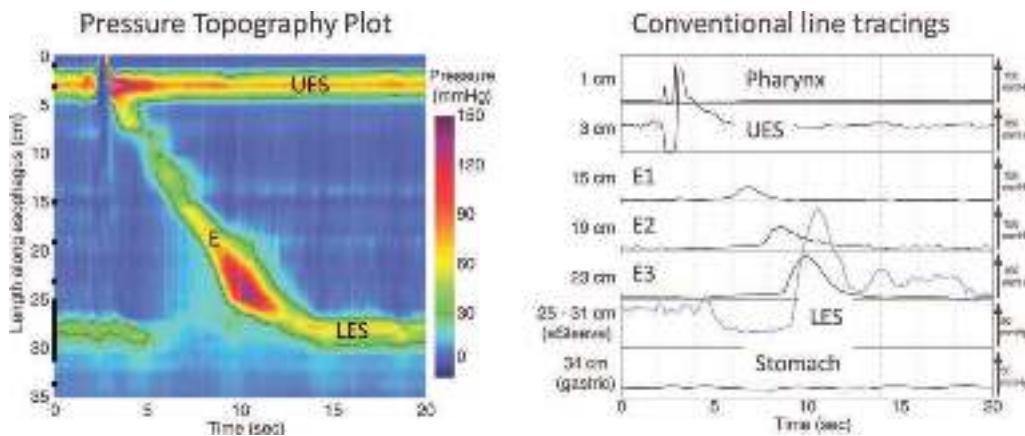


FIGURE 323-1 High-resolution esophageal pressure topography (*right*) and conventional manometry (*left*) of a normal swallow. E, esophageal body; LES, lower esophageal sphincter; UES, upper esophageal sphincter.

atypical symptoms or an inexplicably poor response to therapy. Intraluminal impedance monitoring can be added to pH monitoring to detect reflux events irrespective of whether or not they are acidic, potentially increasing the sensitivity of the study.

STRUCTURAL DISORDERS

■ HIATAL HERNIA

Hiatal hernia is a herniation of viscera, most commonly the stomach, into the mediastinum through the esophageal hiatus of the diaphragm. Four types of hiatal hernia are distinguished, with type I, or sliding hiatal hernia, composing at least 95% of the overall total. A sliding hiatal hernia is one in which the gastroesophageal junction and gastric cardia translocate cephalad as a result of weakening of the phrenoesophageal ligament attaching the gastroesophageal junction to the diaphragm at the hiatus and dilatation of the diaphragmatic hiatus. The incidence of sliding hernia increases with age. True to its name, sliding hernias enlarge with increased intraabdominal pressure, swallowing, and respiration. Conceptually, sliding hernias are the result of wear and tear: increased intraabdominal pressure from abdominal obesity, pregnancy, etc., along with hereditary factors predisposing to the condition. The main significance of sliding hernias is the propensity of affected individuals to have GERD.

Type II, III, and IV hiatal hernias are all subtypes of paraesophageal hernia in which the herniation into the mediastinum includes a visceral structure other than the gastric cardia. With type II and III paraesophageal hernias, the gastric fundus also herniates, with the distinction being that in type II, the gastroesophageal junction remains fixed at the hiatus, whereas type III is a combined sliding and paraesophageal hernia. With type IV hiatal hernias, viscera other than the stomach herniate into the mediastinum, most commonly the colon. With type II and III paraesophageal hernias, the stomach may twist as it herniates, and large paraesophageal hernias can lead to an ‘upside down stomach,’ gastric volvulus, and even strangulation of the stomach. Because of this risk, surgical repair is often advocated for large paraesophageal hernias, particularly when they are symptomatic.

■ RINGS AND WEBS

A lower esophageal mucosal ring, also called a *B ring*, is a thin membranous narrowing at the squamocolumnar mucosal junction (Fig. 323-2). Its origin is unknown, but B rings are demonstrable in ~10–15% of the general population and are usually asymptomatic. When the lumen diameter is <13 mm, distal rings are usually associated with episodic solid food dysphagia and are called *Schatzki rings*. Patients typically present older than 40 years, consistent with an acquired rather than congenital origin. Schatzki ring is one of the most common causes of intermittent food impaction, also known as “steakhouse syndrome” because meat is a typical instigator. Symptomatic rings are readily treated by dilation.

Web-like constrictions higher in the esophagus can be of congenital or inflammatory origin. Asymptomatic cervical esophageal webs are demonstrated in ~10% of people and typically originate along the anterior aspect of the esophagus. Depending on the degree of impingement, they can cause intermittent dysphagia to solids similar to Schatzki rings and are similarly treated with dilation. The combination of symptomatic proximal esophageal webs and iron-deficiency anemia in middle-aged women constitutes Plummer-Vinson or Paterson-Kelly syndrome.

■ DIVERTICULA

Esophageal diverticula are categorized by location, with the most common being epiphrenic, hypopharyngeal (Zenker’s), and midesophageal. Epiphrenic and Zenker’s diverticula are false diverticula involving herniation of the mucosa and submucosa through the muscular layer of the esophagus. These lesions result from increased intraluminal pressure associated with distal obstruction. In the case of Zenker’s, the obstruction is a stenotic cricopharyngeus muscle (upper esophageal sphincter), and the hypopharyngeal herniation most commonly occurs in an area of natural weakness proximal to the cricopharyngeus known

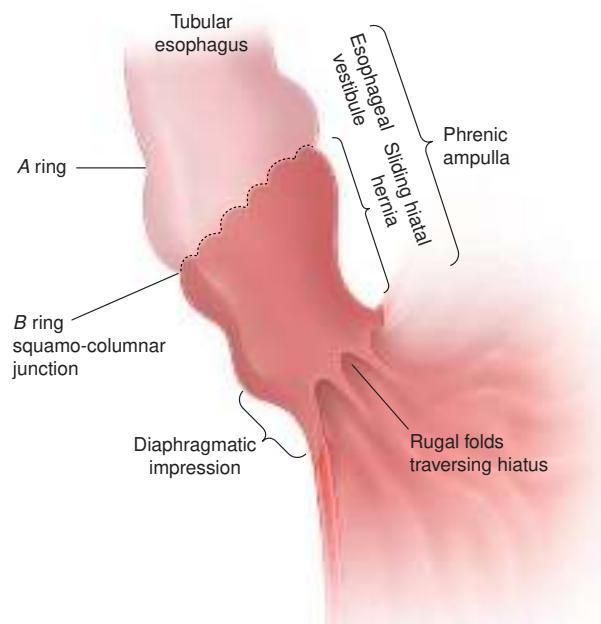


FIGURE 323-2 Radiographic anatomy of the gastroesophageal junction.

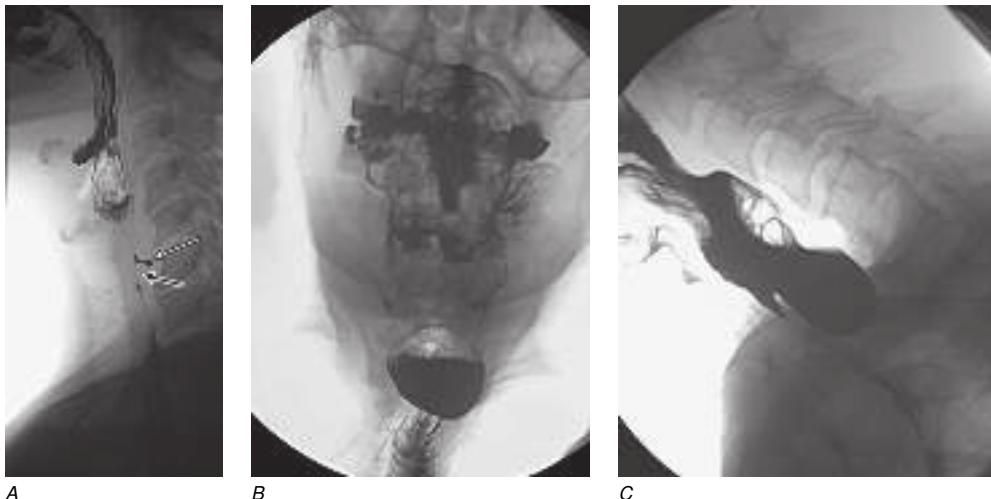


FIGURE 323-3 Examples of small (A) and large (B, C) Zenker's diverticula arising from Killian's triangle in the distal hypopharynx. Smaller diverticula are evident only during the swallow, whereas larger ones retain food and fluid.

as *Killian's triangle* (Fig. 323-3). Small Zenker's diverticula are usually asymptomatic, but when they enlarge sufficiently to retain food and saliva, they can be associated with dysphagia, halitosis, and aspiration. Treatment is by surgical diverticulectomy and cricopharyngeal myotomy or transoral, endoscopic marsupialization.

Epiphrenic diverticula are often associated with achalasia, esophageal hypercontractile disorders, or a distal esophageal stricture. Midesophageal diverticula may be caused by traction from adjacent inflammation (tuberculosis, histoplasmosis), in which case they are true diverticula involving all layers of the esophageal wall, or by pulsion associated with esophageal motor disorders. Midesophageal and epiphrenic diverticula are often asymptomatic until they enlarge sufficiently to retain food and cause dysphagia and regurgitation. Symptoms attributable to the diverticula tend to correlate more with the underlying esophageal disorder than the size of the diverticula. Large diverticula can be removed surgically, usually in conjunction with a myotomy if the underlying motility disorder is identified. Diffuse intramural esophageal pseudodiverticulosis is a rare entity that results from dilatation of the excretory ducts of submucosal esophageal glands (Fig. 323-4). Esophageal candidiasis and proximal esophageal strictures are commonly found in association with this disorder.

TUMORS

Esophageal cancer occurs in ~4.5/100,000 people in the United States, with the associated mortality being 3.9/100,000. It is ~10 times less common than colorectal cancer but kills about one-quarter as many patients. These statistics emphasize both the rarity and lethality of esophageal cancer. One notable trend is the shift of dominant esophageal cancer type from squamous cell to adenocarcinoma, strongly linked to reflux disease and Barrett's metaplasia. Other distinctions between cell types are the predilection for adenocarcinoma to affect the distal esophagus in white males and for squamous cell carcinoma to affect the more proximal esophagus in black males with the added risk factors of smoking, alcohol consumption, caustic injury, and human papillomavirus infection (Chap. 80).

The typical presentation of esophageal cancer is of progressive solid food dysphagia and weight loss. Associated symptoms may include odynophagia, iron deficiency, cough from tracheoesophageal fistula, and hoarseness from left recurrent laryngeal nerve injury. Generally, respiratory symptoms are manifestations of locally invasive or even metastatic disease. Even when detected as a small lesion, esophageal cancer has poor survival because of the abundant esophageal lymphatics leading to regional lymph node metastases.

Benign esophageal tumors are uncommon and usually discovered incidentally. They include gastrointestinal stromal tumors, leiomyoma,

fibrovascular polyps, squamous papilloma, granular cell tumors, lipomas, mesenchymal neoplasms, and inflammatory fibroid polyps.

CONGENITAL ANOMALIES

The most common congenital esophageal anomaly is esophageal atresia, occurring in ~1 in 5000 live births. Atresia can occur in several permutations, the common denominator being developmental failure



FIGURE 323-4 Intramural esophageal pseudodiverticulosis associated with chronic obstruction. Invaginations of contrast into the esophageal wall outline deep esophageal glands.

of fusion between the proximal and distal esophagus associated with a tracheoesophageal fistula, most commonly with the distal segment excluded. Alternatively, there can be an H-type configuration in which esophageal fusion has occurred, but with a tracheoesophageal fistula. Esophageal atresia is usually recognized and corrected surgically within the first few days of life. Later life complications include dysphagia from anastomotic strictures or absent peristalsis and reflux, which can be severe. Less common developmental anomalies include congenital esophageal stenosis, webs, and duplications.

Dysphagia can also result from congenital abnormalities that cause extrinsic compression of the esophagus. In dysphagia lusoria, the esophagus is compressed by an aberrant right subclavian artery arising from the descending aorta and passing behind the esophagus. Alternatively, vascular rings may surround and constrict the esophagus.

Heterotopic gastric mucosa, also known as an esophageal inlet patch, is a focus of gastric-type epithelium in the proximal cervical esophagus; the estimated prevalence is 4–5%. The inlet patch is thought to result from incomplete replacement of embryonic columnar epithelium with squamous epithelium. The majority of inlet patches are asymptomatic, but acid production can occur as most contain fundic-type gastric epithelium with parietal cells.

ESOPHAGEAL MOTILITY DISORDERS

Esophageal motility disorders are diseases attributable to esophageal neuromuscular dysfunction commonly associated with dysphagia, chest pain, or heartburn. The major entities are achalasia, diffuse esophageal spasm (DES), jackhammer esophagus, and GERD. Motility disorders can also be secondary to systemic disease processes, as is the case with pseudoachalasia, Chagas' disease, and scleroderma. Not included in this discussion are diseases affecting the pharynx and proximal esophagus, the impairment of which is almost always part of a more global neuromuscular disease process.

■ ACHALASIA

Achalasia is a rare disease caused by loss of ganglion cells within the esophageal myenteric plexus, with a population incidence estimated to be 1–3 per 100,000 and presentation usually occurring between age 25 and 60 years. With long-standing disease, aganglionosis is noted. The disease involves both excitatory (cholinergic) and inhibitory (nitric oxide) ganglionic neurons. Functionally, inhibitory neurons mediate deglutitive LES relaxation and the sequential propagation of peristalsis. Their absence leads to impaired deglutitive LES relaxation and absent peristalsis. Increasing evidence suggests that the ultimate cause of ganglion cell degeneration in achalasia is an autoimmune process attributable to a latent infection with human herpes simplex virus 1 combined with genetic susceptibility.

Long-standing achalasia is characterized by progressive dilatation and sigmoid deformity of the esophagus with hypertrophy of the LES. Clinical manifestations may include dysphagia, regurgitation, chest pain, and weight loss. Most patients report solid and liquid food dysphagia. Regurgitation occurs when food, fluid, and secretions are retained in the dilated esophagus. Patients with advanced achalasia are at risk for bronchitis, pneumonia, or lung abscess from chronic regurgitation and aspiration. Chest pain may manifest early in the course of achalasia. Patients describe a squeezing, pressure-like retrosternal pain, sometimes radiating to the neck, arms, jaw, and back. Paradoxically, some patients complain of heartburn that may be a chest pain equivalent. Treatment of achalasia is less effective at alleviating chest pain than it is in relieving dysphagia or regurgitation.

The differential diagnosis of achalasia includes jackhammer esophagus, DES, Chagas' disease, opioid-induced esophageal dysmotility, and pseudoachalasia. Chagas' disease is endemic in areas of central Brazil, Venezuela, and northern Argentina and spread by the bite of the reduviid (kissing) bug that transmits the protozoan *Trypanosoma cruzi*. The chronic phase of the disease develops years after infection and results from destruction of autonomic ganglion cells throughout the body, including the heart, gut, urinary tract, and respiratory tract. Manometric features of achalasia have been described in patients on chronic opioids and may be confused with primary achalasia. Tumor

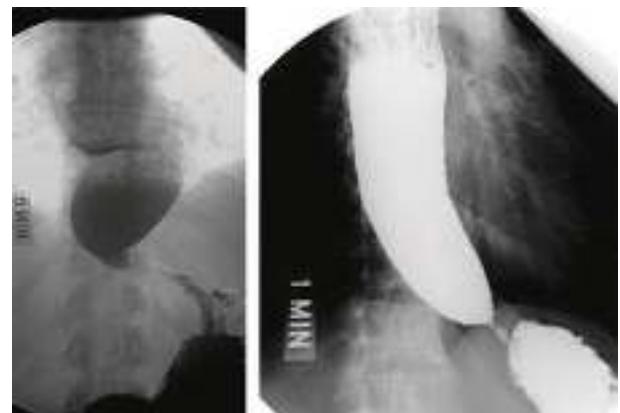


FIGURE 323-5 Achalasia with esophageal dilatation, tapering at the gastroesophageal junction, and an air-fluid level within the esophagus. The example on the *left* shows sigmoid deformity with very advanced disease.

infiltration, most commonly seen with carcinoma in the gastric fundus or distal esophagus, can also mimic primary achalasia. The resultant "pseudoachalasia" accounts for up to 5% of suspected cases and is more likely with advanced age, abrupt onset of symptoms (<1 year), and weight loss. Hence, endoscopy is a necessary part of the evaluation of achalasia. When the clinical suspicion for pseudoachalasia is high and endoscopy nondiagnostic, computed tomography (CT) scanning or EUS may be of value. Rarely, pseudoachalasia can result from a paraneoplastic syndrome with circulating antineuronal antibodies.

Achalasia is diagnosed by barium swallow x-ray and/or esophageal manometry. Endoscopy excludes tumors or benign mechanical strictures of the esophagogastric junction. The barium swallow x-ray appearance is of a dilated esophagus with poor emptying, an air-fluid level, and tapering at the LES giving it a beak-like appearance (Fig. 323-5). Occasionally, an epiphrenic diverticulum is observed. In long-standing achalasia, the esophagus may assume a sigmoid configuration. The diagnostic criteria for achalasia with esophageal manometry are impaired LES relaxation and absent peristalsis. High-resolution manometry has somewhat advanced this diagnosis; three subtypes of achalasia are differentiated based on the pattern of pressurization in the nonperistaltic esophagus (Fig. 323-6). Because manometry identifies early disease before esophageal dilatation and food retention, it is the most sensitive diagnostic test.

No method of preventing or "curing" achalasia is known. Therapy is thus directed at reducing LES pressure so that gravity and esophageal pressurization permit esophageal emptying. While peristalsis does not recover, remnants of peristalsis masked by esophageal pressurization and dilatation prior to therapy may be demonstrable following effective treatment. LES pressure can be reduced by pharmacologic therapy, pneumatic balloon dilation, or LES myotomy by means of submucosal endoscopy or laparoscopic surgery. Pharmacologic therapies are relatively ineffective but can be offered as temporizing therapies. Nitrates or calcium channel blockers are administered before eating but should be used with caution because of their effects on blood pressure. Botulinum toxin, injected into the LES under endoscopic guidance, inhibits acetylcholine release from nerve endings and improves dysphagia in about two-thirds of cases for at least 6 months. Sildenafil and alternative phosphodiesterase inhibitors effectively decrease LES pressure, but practicalities limit their clinical use in achalasia.

The only durable therapies for achalasia are pneumatic dilation and LES myotomy. Pneumatic dilation, with a reported efficacy ranging from 32 to 98%, is an endoscopic technique using a noncompliant, cylindrical balloon dilator positioned across the LES and inflated to a diameter of 3–4 cm. The major complication is perforation, with a reported incidence of 0.5–5%. The most common surgical procedure for achalasia is laparoscopic Heller myotomy, usually performed in conjunction with an antireflux procedure (partial fundoplication); good to excellent results are reported in 62–100% of cases. A European

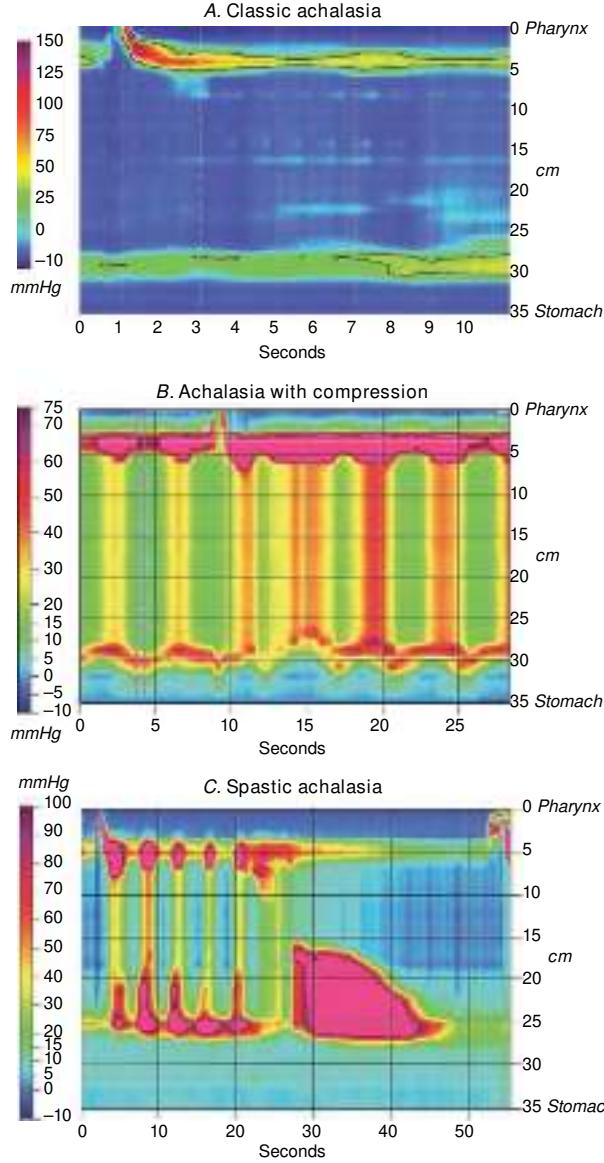


FIGURE 323-6 Three subtypes of achalasia: classic (A), with esophageal compression (B), and spastic achalasia (C) imaged with pressure topography. All are characterized by impaired lower esophageal sphincter (LES) relaxation and absent peristalsis. However, classic achalasia has minimal pressurization of the esophageal body, whereas substantial fluid pressurization is observed in achalasia with esophageal compression, and spastic esophageal contractions are observed with spastic achalasia.

randomized controlled trial demonstrated an equivalent response rate of ~90% for both pneumatic dilation and laparoscopic Heller myotomy at 5-year follow-up. Occasionally, patients with advanced disease fail to respond to pneumatic dilation or Heller myotomy or relapse years after response to primary therapy. In such refractory cases, esophageal resection with gastric pull-up or interposition of a segment of transverse colon may be the only option other than gastrostomy feeding.

An endoscopic approach to LES myotomy is increasingly available, referred to as peroral esophageal myotomy (POEM). This technique involves the creation of a tunnel in the submucosa of the esophageal wall through which the circular muscle of the LES and distal esophagus are transected with electrocautery. As expected, GERD is common after POEM but managed effectively with medications. Potential advantages over the conventional laparoscopic approach include avoidance of surgical disruption of the diaphragmatic hiatus and more rapid

recovery. An international, multicenter, randomized trial of POEM and pneumatic dilation demonstrated greater symptom relief with POEM compared to dilation at 2 years. A European, multicenter, randomized trial of POEM and Heller myotomy reported similar efficacy for symptom relief, with exceeded 80% with either modality.

In untreated or inadequately treated achalasia, esophageal dilatation predisposes to stasis esophagitis. Prolonged stasis esophagitis is the likely explanation for the association between achalasia and esophageal squamous cell cancer. Tumors develop after years of achalasia, usually in the setting of extreme esophageal dilatation, with the overall squamous cell cancer risk increased 17-fold compared to controls.

■ DIFFUSE ESOPHAGEAL SPASM

DES is manifested by episodes of dysphagia and chest pain attributable to abnormal esophageal contractions with normal deglutive LES relaxation. The pathophysiology and natural history of DES are poorly defined. Radiographically, DES has been characterized by tertiary contractions or a "corkscrew esophagus" (Fig. 323-7), but in many instances, these abnormalities are indicative of achalasia. Manometrically, a variety of defining features have been proposed including uncoordinated ("spastic") activity in the distal esophagus, spontaneous and repetitive contractions, or high-amplitude and prolonged contractions. High-resolution manometry has defined DES by the occurrence of contractions in the distal esophagus with short latency relative to the time of the pharyngeal contraction, a dysfunction indicative of impairment of inhibitory myenteric plexus neurons. When defined with this restrictive criterion (Fig. 323-8), DES is substantially less common than achalasia.

Esophageal chest pain closely mimics angina pectoris. Features suggesting esophageal pain include pain that is nonexertional, prolonged, meal-related, relieved with antacids, and accompanied by heartburn, dysphagia, or regurgitation and interrupts sleep. However, all of these features exhibit overlap with cardiac pain, which still must be the primary consideration. Furthermore, even within the spectrum of esophageal diseases, both chest pain and dysphagia are also characteristic of peptic or infectious esophagitis. Only after these more common entities have been excluded by evaluation and/or treatment should a diagnosis of DES be pursued.

Although DES is diagnosed by manometry, endoscopy is useful to identify alternative structural and inflammatory lesions that may cause chest pain. Radiographically, a "corkscrew esophagus," "rosary bead esophagus," pseudodiverticula, or curling can be indicative of DES, but these are also found with spastic achalasia. Given these vagaries of defining DES and the resultant heterogeneity of patients identified for

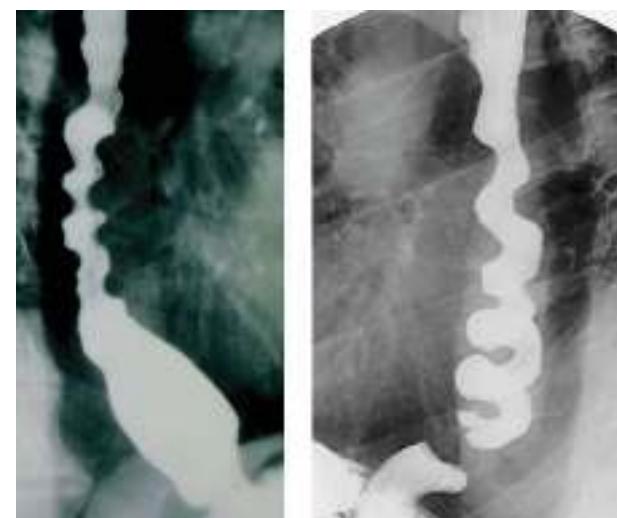


FIGURE 323-7 Diffuse esophageal spasm. The characteristic "corkscrew" esophagus results from spastic contraction of the circular muscle in the esophageal wall; more precisely, this is actually a helical array of muscle. These findings are also seen with spastic achalasia.

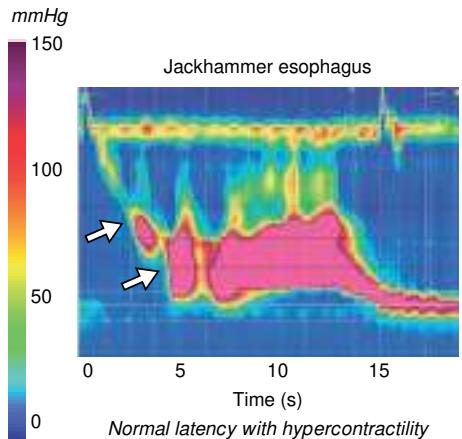


FIGURE 323-8 Esophageal pressure topography of the two major variants of esophageal spasm: jackhammer esophagus (*left*) and diffuse esophageal spasm (*right*). Jackhammer esophagus is defined by the extraordinarily vigorous and repetitive contractions with normal peristaltic onset and normal latency of the contraction. Diffuse esophageal spasm is similar but primarily defined by a short latency (premature) contraction.

inclusion in therapeutic trials, it is not surprising that trial results have been disappointing. Only small, uncontrolled trials exist, reporting response to nitrates, calcium channel blockers, hydralazine, botulinum toxin, and anxiolytics. POEM with distal esophageal myotomy or surgical myotomy should be considered only with severe weight loss or intractable pain. These indications are extremely rare.

■ NONSPECIFIC MANOMETRIC FINDINGS

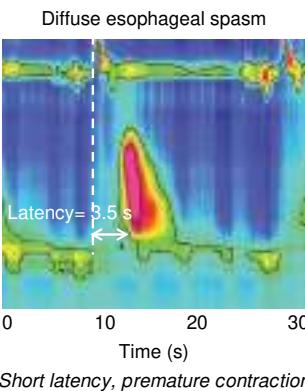
Manometric studies done to evaluate chest pain and/or dysphagia often report minor abnormalities (e.g., hypertensive or hypotensive peristalsis, hypertensive LES) that are insufficient to diagnose either achalasia or DES. These findings are of unclear significance. Reflux and psychiatric diagnoses, particularly anxiety and depression, are common among such individuals. A lower visceral pain threshold and symptoms of irritable bowel syndrome are noted in more than half of such patients. Consequently, therapy for these individuals should target either the most common esophageal disorder, GERD, or cognitive disorders that may be present.

GASTROESOPHAGEAL REFLUX DISEASE

The current conception of GERD is that it encompasses a family of conditions with the commonality that they are caused by gastroesophageal reflux resulting in either troublesome symptoms or an array of potential esophageal and extraesophageal manifestations. It is estimated that 10–15% of adults in the United States are affected by GERD, although such estimates are based on population studies of self-reported chronic heartburn. With respect to the esophagus, the spectrum of injury includes esophagitis, stricture, Barrett's esophagus, and adenocarcinoma (Fig. 323-9). Of particular concern is the rising incidence of esophageal adenocarcinoma, an epidemiologic trend that parallels the increasing incidence of GERD. About 9200 incident cases of esophageal adenocarcinoma were noted in the United States in 2020 (estimated as half of all esophageal cancers); this disease burden has increased two- to sixfold in the past 20 years.

■ PATHOPHYSIOLOGY

The best-defined subset of GERD patients, albeit a minority overall, have esophagitis. Esophagitis occurs when refluxed gastric acid and pepsin induce inflammation of the esophageal mucosa that leads to microscopic injury and macroscopic erosions and ulcers. Experimental evidence supports a cytokine-mediated inflammatory pathway rather than direct caustic injury to the esophageal epithelium. Note that some degree of gastroesophageal reflux is normal, physiologically intertwined with the mechanism of belching (transient LES relaxation), but esophagitis results from excessive reflux, often accompanied by impaired clearance of the refluxed gastric juice. Restricting reflux



to that which is physiologically intended depends on the anatomic and physiologic integrity of the esophagogastric junction, a complex sphincter comprised of both the LES and the surrounding crural diaphragm. Three dominant mechanisms of esophagogastric junction incompetence are recognized: (1) transient LES relaxations (a vagovagal reflex in which LES relaxation is elicited by gastric distention), (2) LES hypotension, or (3) anatomic distortion of the esophagogastric junction inclusive of hiatal hernia. Of note, the third factor, esophagogastric junction anatomic disruption, is significant both unto itself and also because it interacts with the first two mechanisms. Transient LES relaxations account for ~90% of reflux in normal subjects or GERD patients without hiatal hernia, but patients with hiatal hernia have a more heterogeneous mechanistic profile. Factors tending to exacerbate reflux regardless of mechanism are abdominal obesity, preg-

nancy, gastric hypersecretory states, delayed gastric emptying, disruption of esophageal peristalsis, and gluttony.

After acid reflux, peristalsis returns the refluxed fluid to the stomach, and acid clearance is completed by titration of the residual acid by bicarbonate contained in swallowed saliva. Consequently, two causes of prolonged acid clearance are impaired peristalsis and reduced salivation. Impaired peristaltic emptying can be attributable to disrupted peristalsis or superimposed reflux associated with a hiatal hernia. With superimposed reflux, fluid retained within a sliding hiatal hernia refluxes back into the esophagus during swallow-related LES relaxation, a phenomenon that does not normally occur.

Inherent in the pathophysiologic model of GERD is that gastric juice is harmful to the esophageal epithelium. However, gastric acid hypersecretion is usually not a dominant factor in the development of esophagitis. An obvious exception is with Zollinger-Ellison syndrome, which is associated with severe esophagitis in ~50% of patients. Another caveat is with chronic *Helicobacter pylori* gastritis, which may have a protective effect by inducing atrophic gastritis with concomitant hypoacidity. Pepsin, bile, and pancreatic enzymes within gastric secretions can also injure the esophageal epithelium, but their noxious properties are either lessened without an acidic environment or dependent on acidity for activation. Bile warrants attention because it persists in refluxate despite acid-suppressing medications. Bile can transverse the cell membrane, imparting severe cellular injury in a weakly acidic environment, and has also been invoked as a cofactor in the pathogenesis of Barrett's metaplasia and adenocarcinoma. Hence, the causticity of gastric reflux extends beyond hydrochloric acid.

■ SYMPTOMS

Heartburn and regurgitation are the typical symptoms of GERD. Somewhat less common are dysphagia and chest pain. In each case, multiple potential mechanisms for symptom genesis operate that extend beyond the basic concepts of mucosal erosion and activation of afferent sensory nerves. Specifically, visceral sensitivity is increasingly recognized as a cofactor. Nonetheless, the dominant clinical strategy is empirical treatment with acid inhibitors, reserving further evaluation for those who fail to respond. Important exceptions to this are patients with chest pain or persistent dysphagia, each of which may be indicative of more morbid consequences of GERD or alternative diagnoses. With chest pain, cardiac disease must be carefully considered. In the case of dysphagia, chronic reflux can lead to the development of a peptic stricture, eosinophilic esophagitis (EoE), or adenocarcinoma, each of which benefits from early detection and/or specific therapy.

Extraesophageal syndromes with an established association to GERD include chronic cough, laryngitis, asthma, and dental erosions. A multitude of other conditions including pharyngitis, chronic



A Erosive esophagitis



B Esophageal stricture with chronic erosive esophagitis



C Barrett's esophagus



D Esophageal adenocarcinoma with Barrett's esophagus

FIGURE 323-9 Endoscopic appearance of (A) peptic esophagitis, (B) a peptic stricture, (C) Barrett's metaplasia, and (D) adenocarcinoma developing within an area of Barrett's esophagus.

bronchitis, pulmonary fibrosis, chronic sinusitis, cardiac arrhythmias, sleep apnea, and recurrent aspiration pneumonia have proposed associations with GERD. However, in both cases, it is important to emphasize the word *association* as opposed to *causation*. In many instances, the disorders likely coexist because of shared pathogenetic mechanisms rather than strict causality. Potential mechanisms for extraesophageal GERD manifestations are either regurgitation with direct contact between the refluxate and supraesophageal structures or via a vagovagal reflex wherein reflux activation of esophageal afferent nerves triggers efferent vagal reflexes such as bronchospasm, cough, or arrhythmias.

■ DIFFERENTIAL DIAGNOSIS

Although generally quite characteristic, symptoms from GERD need to be distinguished from symptoms related to infectious or pill esophagitis, EoE, peptic ulcer disease, dyspepsia, biliary colic, coronary artery disease, and esophageal motility disorders. It is especially important that coronary artery disease be given early consideration because of its potentially lethal implications. The remaining elements of the differential diagnosis can be addressed by endoscopy, upper gastrointestinal series, or esophageal manometry as appropriate. Erosive esophagitis at the esophagogastric junction is the endoscopic hallmark of GERD but identified in only about one-third of patients with GERD. The distinction among etiologies of esophagitis is readily made by endoscopic appearance, but mucosal biopsies may be helpful to evaluate for infectious or eosinophilic inflammation. In terms of endoscopic appearance, the ulcerations seen in peptic esophagitis are usually few and distal, whereas infectious ulcerations are numerous, punctate, and diffuse. EoE characteristically exhibits multiple esophageal rings, linear furrows, white punctate exudate, and strictures. Esophageal ulcerations from pill esophagitis are usually singular and deep at points of luminal narrowing, especially near the carina, with sparing of the distal esophagus.

■ COMPLICATIONS

The complications of GERD are related to chronic esophagitis (bleeding and stricture) and the relationship between GERD and esophageal adenocarcinoma. However, both erosive esophagitis and peptic strictures have become increasingly rare in the era of potent antisecretory medications. Conversely, the most severe histologic consequence of GERD is Barrett's metaplasia with the associated risk of esophageal adenocarcinoma, and the incidence of these lesions has increased, not decreased, in the era of potent acid suppression. Barrett's metaplasia, recognized endoscopically by salmon-colored mucosa extending proximally from the gastroesophageal junction (Fig. 323-9) or histopathologically by the finding of specialized columnar metaplasia, is associated with a significantly increased risk for development of esophageal adenocarcinoma.

Barrett's metaplasia can progress to adenocarcinoma through the intermediate stages of low- and high-grade dysplasia (Fig. 323-10). Owing to this risk, areas of Barrett's metaplasia and especially any included areas of mucosal irregularity should be carefully inspected and extensively biopsied. The rate of cancer development is estimated at 0.1–0.3% per year, but vagaries in definitional criteria and of the extent of Barrett's metaplasia requisite to establish the diagnosis have contributed to variability and inconsistency in this risk assessment. The group at greatest risk is obese white males in their sixth decade of life. Despite common practice, however, the utility of endoscopic screening and surveillance programs intended to control the adenocarcinoma risk remains an open question. Also of note, although in a large, randomized, controlled chemoprevention trial in

Barrett's patients high-dose proton pump inhibitor therapy along with aspirin did significantly better at achieving the primary composite endpoint of delaying all-cause mortality, development of esophageal adenocarcinoma, or progression to high-grade dysplasia, the effect was driven mainly by improved overall survival rather than reduced Barrett's progression or esophageal adenocarcinoma development.

Although the management of Barrett's esophagus remains controversial, the finding of dysplasia in Barrett's, particularly high-grade dysplasia, mandates further intervention. In addition to the high rate of progression to adenocarcinoma, there is also a high prevalence of unrecognized coexisting cancer with high-grade dysplasia. Treatment recommendations for Barrett's esophagus with high-grade dysplasia have evolved over the past several years. Historically, esophagectomy was the gold standard treatment for high-grade dysplasia. However, esophagectomy has a mortality ranging from 3 to 10%, along with substantial morbidity. Prospective studies have demonstrated the efficacy of endoscopic mucosal ablation therapy with substantially less morbidity and essentially no mortality. Consequently, current societal guidelines endorse endoscopic mucosal ablation therapies for the management of high-grade dysplasia.

TREATMENT

Gastroesophageal Reflux Disease

Lifestyle modifications are routinely advocated as GERD therapy. Broadly speaking, these fall into three categories: (1) avoidance of foods that reduce LES pressure, making them "refluxogenic" (these commonly include fatty foods, alcohol, spearmint, peppermint, and possibly coffee and tea); (2) avoidance of acidic foods that are inherently irritating (citrus fruits, tomato-based foods); and (3) adoption of behaviors to minimize reflux and/or heartburn. In general, minimal evidence supports the efficacy of these measures.

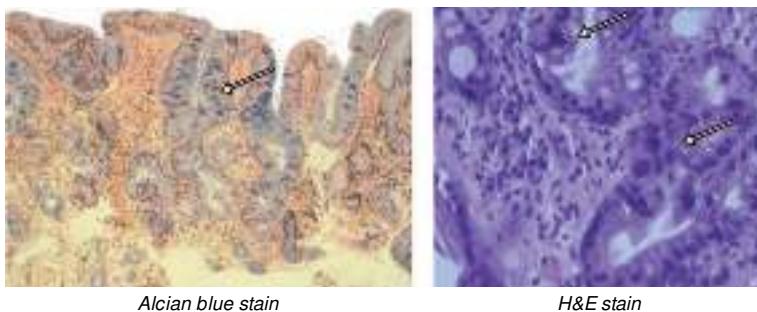


FIGURE 323-10 Histopathology of Barrett's metaplasia and Barrett's metaplasia with high-grade dysplasia. H&E, hematoxylin and eosin.

However, clinical experience dictates that subsets of patients benefit from specific recommendations based on their individual history and symptom profile. A patient with sleep disturbance from nighttime heartburn is more likely to benefit from elevation of the head of the bed and avoidance of eating before retiring. The most broadly applicable recommendation is for weight reduction. Even though the benefit with respect to reflux cannot be assured, the strong epidemiologic relationship between body mass index and GERD and the secondary health gains of weight reduction is beyond dispute.

The dominant pharmacologic approach to GERD management is with inhibitors of gastric acid secretion, and abundant data support the effectiveness of this approach. Pharmacologically reducing the acidity of gastric juice does not prevent reflux, but it ameliorates reflux symptoms and allows esophagitis to heal. The hierarchy of effectiveness among pharmaceuticals parallels their antisecretory potency. Proton pump inhibitors (PPIs) are more efficacious than histamine-2 receptor antagonists (H₂RAs), and both are superior to placebo. No major differences exist among PPIs, and only modest gain is achieved by increased dosage.

Paradoxically, the perceived frequency and severity of heartburn correlate poorly with the presence or severity of esophagitis. When GERD treatments are assessed in terms of resolving heartburn, both efficacy and differences among pharmaceuticals are less clear-cut than with the objective of healing esophagitis. Although the same overall hierarchy of effectiveness exists, observed efficacy rates are lower and vary widely, likely reflecting patient heterogeneity.

Reflux symptoms tend to be chronic, irrespective of esophagitis. Thus, a common management strategy is indefinite treatment with PPIs or H₂RAs as necessary for symptom control. The side effects of PPI therapy are generally minimal. Rare cases of interstitial nephritis and severe, reversible hypomagnesemia have been reported. Vitamin B₁₂ and iron absorption may be compromised and susceptibility to enteric infections, particularly *Clostridium difficile* colitis, increased with treatment. Observational data have also noted an association between PPI exposure and renal disease, dementia, and cardiovascular disease, but the hazard ratios reported in these studies were small, and the potential for unrecognized residual confounding bias was substantial. Population studies have also suggested a slight increased risk of bone fracture with chronic PPI use suggesting an impairment of calcium absorption, but prospective studies have failed to corroborate this. Nonetheless, as with any medication, PPI dosage should be minimized to that necessary for the clinical indication.

Laparoscopic Nissen fundoplication, wherein the proximal stomach is wrapped around the distal esophagus to create an antireflux barrier, is a surgical alternative to the management of chronic GERD. Just as with PPI therapy, evidence on the utility of fundoplication is strongest for treating esophagitis, and controlled trials suggest similar efficacy to PPI therapy. However, the benefits of fundoplication must be weighed against potential deleterious effects, including surgical morbidity and mortality, postoperative

dysphagia, failure or breakdown requiring reoperation, an inability to belch, and increased bloating, flatulence, and bowel symptoms after surgery.

■ EOSINOPHILIC ESOPHAGITIS

EoE is increasingly recognized in adults and children around the world. Current prevalence estimates in the United States identified 4–8 cases per 10,000 with a predilection for white males between 30 and 40 years of age. The increasing prevalence of EoE is attributable to a combination of an increasing incidence and a growing recognition of the condition. There is also an incompletely understood, but important, interaction between EoE and GERD that may confound the diagnosis of the disease. Genome-wide analysis studied demonstrated susceptibility elements at 5q22 (thymic stromal lymphopoietin) and 2p23 (CAPN14) in EoE.

EoE is diagnosed based on the combination of esophageal symptoms and esophageal mucosal biopsies demonstrating eosinophil-predominant inflammation. Alternative etiologies of esophageal eosinophilia include GERD, drug hypersensitivity, connective tissue disorders, hypereosinophilic syndrome, Crohn's disease, eosinophilic gastroenteritis, and infection. EoE is an immunologic disorder induced by antigen sensitization in susceptible individuals. Food allergens are the dominant triggers, although aeroallergens may also contribute. The natural history of EoE is incompletely understood, but an increased risk of esophageal stricture development paralleling the duration of untreated disease has been noted.

EoE should be strongly considered in children and adults with dysphagia and esophageal food impactions. In preadolescent children, symptom presentations of EoE include chest or abdominal pain, nausea, vomiting, and food aversion. Other symptoms in adults may include atypical chest pain and heartburn. An atopic history of IgE-mediated food allergy, asthma, eczema, and/or allergic rhinitis is present in the majority of patients. Peripheral blood eosinophilia is demonstrable in 25–50% of patients, but the specificity of this finding is problematic in the setting of concomitant atopy. The characteristic endoscopically identified esophageal findings include loss of vascular markings (edema), multiple esophageal rings, longitudinally oriented furrows, and whitish exudate (Fig. 323-11). Histologic confirmation is made with the demonstration of esophageal mucosal eosinophilia (peak density ≥ 15 eosinophils per high-power field) (Fig. 323-12). Complications of EoE include food impaction, esophageal stricture, narrow-caliber esophagus, and rarely esophageal perforation.

The goals of EoE management are symptom control and the prevention of complications. Primary therapy often starts with PPI therapy, which is effective at improving eosinophilic inflammation in 30–50% of patients. Additional first-line therapies include elimination diets or swallowed topical glucocorticoids. Elemental formula diets devoid of allergenic protein are a highly effective therapy but are limited by palatability. Notably, allergy testing by means of either serum IgE or skin prick testing has demonstrated poor sensitivity and specificity in

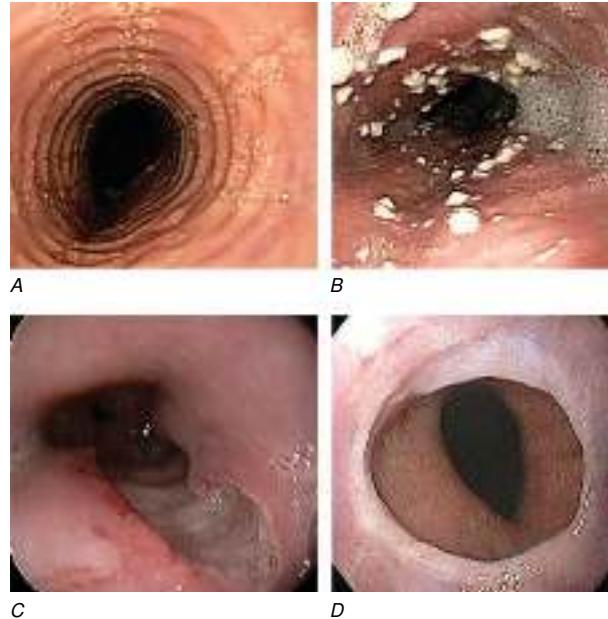


FIGURE 323-11 Endoscopic features of (A) eosinophilic esophagitis (EoE), (B) *Candida* esophagitis, (C) giant ulcer associated with HIV, and (D) a Schatzki ring.

a day for 7–10 days) can be used for immunocompetent hosts, although the disease is typically self-limited after a 1- to 2-week period in such patients. Immunocompromised patients are treated with acyclovir (400 mg orally five times a day for 14–21 days), famciclovir (500 mg orally three times a day), or valacyclovir (1 g orally three times a day).

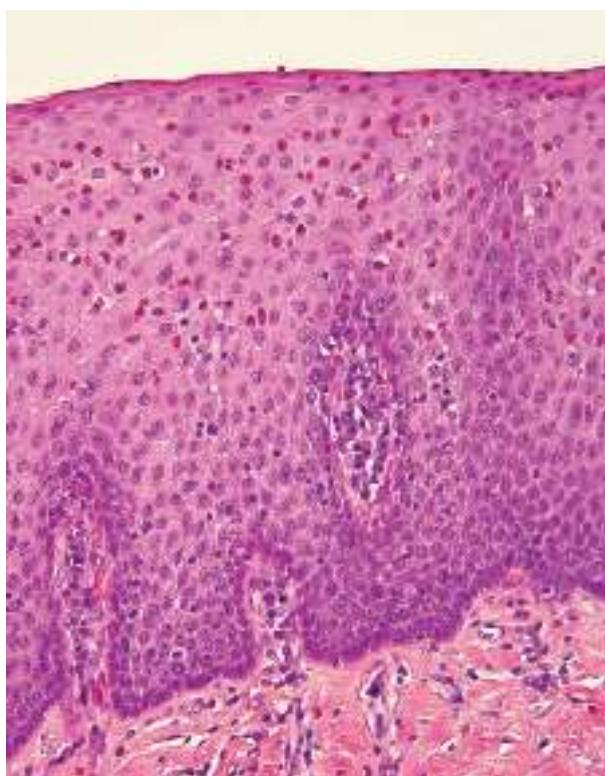


FIGURE 323-12 Histopathology of eosinophilic esophagitis (EoE) showing infiltration of the esophageal squamous epithelium with eosinophils. Additional features of basal cell hyperplasia and lamina propria fibrosis are present. Eosinophilic inflammation can also be seen with gastroesophageal reflux disease.

the identification of foods responsible for EoE in an individual patient. Empiric elimination of common food allergies (milk, wheat, egg, soy, nuts, and seafood) followed by systematic reintroduction has been an effective diet therapy in both children and adults with EoE. The intent of the elimination diet approach is the identification of specific food trigger(s). Swallowed, topical glucocorticoids (e.g., fluticasone propionate or budesonide) are effective in 50–80% of patients, but recurrence of disease is common following the cessation of short-term therapy. Systemic glucocorticoids are not generally recommended due to side effects and lack of proven benefit beyond that achieved with topical glucocorticoids. Biologic therapies targeting allergic cytokine mediators including interleukin (IL) 4, IL-5, and IL-13 have shown promise in initial clinical trials. Esophageal dilation is highly effective at relieving dysphagia in patients with fibrostenosis but does not address the underlying inflammatory process. Dilation should be approached conservatively because of the risk of deep, esophageal mural laceration or perforation in the stiff-walled esophagus that is characteristic of the disease.

INFECTIOUS ESOPHAGITIS

As a result of the increased use of immunosuppression for organ transplantation and chronic inflammatory diseases and use of chemotherapy agents, along with the AIDS epidemic, infections with *Candida* species, herpesvirus, and cytomegalovirus (CMV) have become relatively common. Although rare, infectious esophagitis also occurs among the non-immunocompromised, with herpes simplex and *Candida albicans* being the most common pathogens. Among AIDS patients, infectious esophagitis becomes more common as the CD4 count declines; cases are rare with a CD4 count >200 and common when <100. HIV itself may also be associated with a self-limited syndrome of acute esophageal ulceration with oral ulcers and a maculopapular skin rash at the time of seroconversion. Additionally, some patients with advanced disease have deep, persistent esophageal ulcers treated with oral glucocorticoids or thalidomide. However, with the widespread use of highly effective antiviral therapies, a reduction in these HIV complications has been noted.

Regardless of the infectious agent, odynophagia is a characteristic symptom of infectious esophagitis; dysphagia, chest pain, and hemorrhage are also common. Odynophagia is uncommon with reflux esophagitis, so its presence should always raise suspicion of an alternative etiology.

CANDIDA ESOPHAGITIS

Candida is normally found in the throat but can become pathogenic and produce esophagitis in a compromised host; *C. albicans* is most common. *Candida* esophagitis also occurs with esophageal stasis secondary to esophageal motor disorders and diverticula. Patients complain of odynophagia and dysphagia. If oral thrush is present, empirical therapy is appropriate, but co-infection is common, and persistent symptoms should lead to prompt endoscopy with biopsy, which is the most useful diagnostic evaluation. *Candida* esophagitis has a characteristic appearance of white plaques or exudate with friability. Oral fluconazole (400 mg on the first day, followed by 200 mg daily) for 14–21 days is the preferred treatment. Patients refractory to fluconazole may respond to voriconazole or posaconazole. Alternatively, poorly responsive patients or those who cannot swallow medications can be treated with an intravenous echinocandin.

HERPETIC ESOPHAGITIS

Herpes simplex virus type 1 or 2 may cause esophagitis. Vesicles on the nose and lips may coexist and are suggestive of a herpetic etiology. Varicella-zoster virus can also cause esophagitis in children with chickenpox or adults with zoster. The characteristic endoscopic findings are vesicles and small, superficial ulcerations. Because herpes simplex infections are limited to squamous epithelium, biopsies from the ulcer margins are most likely to reveal the characteristic ground-glass nuclei, eosinophilic Cowdry's type A inclusion bodies, and giant cells. Culture or polymerase chain reaction (PCR) assays are helpful to identify acyclovir-resistant strains. Acyclovir (200 mg orally five times

In patients with severe odynophagia, intravenous acyclovir, 5 mg/kg every 8 h for 7–14 days, reduces this morbidity.

■ CYTOMEGALOVIRUS

CMV esophagitis occurs primarily in immunocompromised patients, particularly those with HIV, patients with malignancy, and recipients of bone marrow or organ transplants. CMV is usually activated from a latent stage. Endoscopically, CMV lesions appear as large serpiginous ulcers in an otherwise normal mucosa, particularly in the distal esophagus. Biopsies from the ulcer bases have the greatest diagnostic yield for finding the pathognomonic large nuclear or cytoplasmic inclusion bodies. Immunohistology with monoclonal antibodies to CMV and *in situ* hybridization tests are useful for early diagnosis. Data on therapy for CMV esophagitis are limited. Treatment studies of CMV gastrointestinal disease have demonstrated effectiveness of both ganciclovir (5 mg/kg every 12 h IV) and valganciclovir (900 mg orally every 12 h). Therapy is continued until healing, which may take 3–6 weeks. Maintenance therapy may be needed for patients with relapsing disease.

MECHANICAL TRAUMA AND IATROGENIC INJURY

ESOPHAGEAL PERFORATION

Most cases of esophageal perforation are from instrumentation of the esophagus or trauma. Alternatively, forceful vomiting or retching can lead to spontaneous rupture at the gastroesophageal junction (Boerhaave's syndrome). More rarely, corrosive esophagitis or neoplasms lead to perforation. Instrument perforation from endoscopy or nasogastric tube placement typically occurs in the hypopharynx or at the gastroesophageal junction. Perforation may also occur at the site of a stricture in the setting of endoscopic food disimpaction or esophageal dilation. Esophageal perforation causes pleuritic retrosternal pain that can be associated with pneumomediastinum and subcutaneous emphysema. Mediastinitis is a major complication of esophageal perforation, and prompt recognition is key to optimizing outcome. CT of the chest is most sensitive in detecting mediastinal air. Esophageal perforation is confirmed by a contrast swallow, usually Gastrografin followed by thin barium. Treatment includes nasogastric suction and parenteral broad-spectrum antibiotics with prompt surgical drainage and repair in noncontained leaks. Conservative therapy with NPO status and antibiotics without surgery may be appropriate in cases of contained perforation that are detected early. Endoscopic clipping or stent placement may be indicated in nonoperated iatrogenic perforations or nonoperable cases such as perforated tumors.

■ MALLORY WEISS TEAR

Vomiting, retching, or vigorous coughing can cause a nontransmural tear at the gastroesophageal junction that is a common cause of upper gastrointestinal bleeding. Most patients present with hematemesis. Antecedent vomiting is the norm but not always evident. Bleeding usually abates spontaneously, but protracted bleeding may respond to local epinephrine or cauterization therapy, endoscopic clipping, or angiographic embolization. Surgery is rarely needed.

■ RADIATION ESOPHAGITIS

Radiation esophagitis can complicate treatment for thoracic cancers, especially breast and lung cancers, with the risk proportional to radiation dosage. Radiosensitizing drugs such as doxorubicin, bleomycin, cyclophosphamide, and cisplatin also increase the risk. Dysphagia and odynophagia may last weeks to months after therapy. The esophageal mucosa becomes erythematous, edematous, and friable. Submucosal fibrosis and degenerative tissue changes and stricturing may occur years after the radiation exposure. Radiation exposure in excess of 5000 cGy has been associated with increased risk of esophageal stricture. Treatment for acute radiation esophagitis is supportive. Chronic strictures are managed with esophageal dilation.

■ CORROSIVE ESOPHAGITIS

Caustic esophageal injury from ingestion of alkali or, less commonly, acid can be accidental or from attempted suicide. Absence of oral

injury does not exclude possible esophageal involvement. Thus, early endoscopic evaluation is recommended to assess and grade the injury to the esophageal mucosa. Severe corrosive injury may lead to esophageal perforation, bleeding, stricture, and death. Glucocorticoids have not been shown to improve the clinical outcome of acute corrosive esophagitis and are not recommended. Healing of more severe grades of caustic injury is commonly associated with severe stricture formation and often requires repeated dilation.

■ PILL ESOPHAGITIS

Pill-induced esophagitis occurs when a swallowed pill fails to traverse the entire esophagus and lodges within the lumen. Generally, this is attributed to poor "pill-taking habits": inadequate liquid with the pill or lying down immediately after taking a pill. The most common location for the pill to lodge is in the mid-esophagus near the crossing of the aorta or carina. Extrinsic compression from these structures halts the movement of the pill or capsule. Since initially reported in 1970, >1000 cases of pill esophagitis have been reported, suggesting that this is not an unusual occurrence. A wide variety of medications are implicated, with the most common being doxycycline, tetracycline, quinidine, phenytoin, potassium chloride, ferrous sulfate, nonsteroidal anti-inflammatory drugs (NSAIDs), and bisphosphonates.

Typical symptoms of pill esophagitis are the sudden onset of chest pain and odynophagia. Characteristically, the pain will develop over a period of hours or will awaken the individual from sleep. A classic history in the setting of ingestion of recognized pill offenders obviates the need for diagnostic testing in most patients. When endoscopy is performed, localized ulceration or inflammation is evident. Histologically, acute inflammation is typical. Chest CT imaging will sometimes reveal esophageal thickening consistent with transmural inflammation. Although the condition usually resolves within days to weeks, symptoms may persist for months and stricture can develop in severe cases. No specific therapy is known to hasten the healing process, but antisecretory medications are frequently prescribed to remove concomitant reflux as an aggravating factor. When healing results in stricture formation, dilation is indicated.

■ FOREIGN BODIES AND FOOD IMPACTION

Food or foreign bodies may lodge in the esophagus, causing complete obstruction, which in turn can cause an inability to handle secretions (foaming at the mouth) and severe chest pain. Food impaction may occur due to peptic stricture, carcinoma, Schatzki ring, EoE, or simply inattentive eating. If it does not spontaneously resolve, impacted food should be removed endoscopically. Use of meat tenderizer enzymes to facilitate passage of a meat bolus is discouraged because of potential esophageal injury. Glucagon (1 mg IV) is sometimes tried before endoscopic dislodgement. After emergent treatment, patients should be evaluated for potential causes of the impaction with treatment rendered as indicated.

ESOPHAGEAL MANIFESTATIONS OF SYSTEMIC DISEASE

■ SCLERODERMA AND CONNECTIVE TISSUE DISORDERS

Scleroderma esophagus (hypotensive LES and absent esophageal contractility) was initially described as a manifestation of scleroderma or other collagen vascular diseases and thought to be specific for these disorders. However, this nomenclature subsequently has been discarded because an estimated half of qualifying patients do not have an identifiable rheumatologic disease, and reflux disease is often the only identifiable association. When scleroderma esophagus occurs as a manifestation of a connective tissue disorder, the histopathologic findings are of infiltration and destruction of the esophageal muscularis propria with collagen deposition and fibrosis and reduction in the number of interstitial cells of Cajal. The pathogenesis of absent peristalsis and LES hypotension in the absence of a connective tissue disorder is unknown. Regardless of the underlying cause, the manometric abnormalities predispose patients to severe GERD due to inadequate LES barrier function combined with poor esophageal clearance

of refluxed acid. Dysphagia may also be manifest but is generally mild and alleviated by eating in an upright position and using liquids to facilitate solid emptying.

■ DERMATOLOGIC DISEASES

A host of dermatologic disorders (lichen planus, pemphigus vulgaris, bullous pemphigoid, cicatricial pemphigoid, Behcet's syndrome, and epidermolysis bullosa) can affect the oropharynx and esophagus, particularly the proximal esophagus, with blisters, bullae, ulceration, webs, and strictures. Topical or systemic anti-inflammatory therapy is effective for mucosal healing. Stevens-Johnson syndrome and graft-versus-host disease can also involve the esophagus. Esophageal dilation may be necessary to treat strictures.

■ FURTHER READING

- F GT, K DA: Eosinophilic esophagitis. *N Engl J Med* 373:1640, 2015.
 H I et al: American Gastroenterological Institute and the joint task force on allergy-immunology practice parameters clinical guidelines for the management of eosinophilic esophagitis. *Gastroenterology* 158:1776, 2020.
 K PI, B G: The spectrum of achalasia: Lessons from studies of pathophysiology and high-resolution manometry. *Gastroenterology* 145:954, 2013.
 K PJ et al: American Gastroenterological Association Institute technical review on the management of gastroesophageal reflux disease. *Gastroenterology* 135:1392, 2008.
 K DA et al: Phenotypes of gastroesophageal reflux disease: where Rome, Lyon, and Montreal meet. *Clin Gastroenterol Hepatol* 18:767, 2020.
 P JE, G AJ: Achalasia: A systematic review. *JAMA* 313:1841, 2015.
 S NJ et al: Diagnosis and management of Barrett's esophagus. *Am J Gastroenterol* 111:30, 2016.
 S SJ, S RF: Barrett's esophagus. *N Engl J Med* 371:836, 2014.

324

Peptic Ulcer Disease and Related Disorders

John Del Valle

PEPTIC ULCER DISEASE

A *peptic ulcer* is defined as disruption of the mucosal integrity of the stomach and/or duodenum leading to a local defect or excavation due to active inflammation. Although burning epigastric pain exacerbated by fasting and improved with meals is a symptom complex associated with peptic ulcer disease (PUD), it is now clear that >90% patients with this symptom complex (dyspepsia) do not have ulcers and that the majority of patients with peptic ulcers may be asymptomatic. Ulcers occur within the stomach and/or duodenum and are often chronic in nature. Acid peptic disorders are very common in the United States, with 4 million individuals (new cases and recurrences) affected per year. Lifetime overall prevalence of PUD in the United States is ~8.4% with a slightly higher prevalence in men. PUD significantly affects quality of life by impairing overall patient well-being and contributing substantially to work absenteeism. Moreover, an estimated 15,000 deaths per year occur as a consequence of complicated PUD. The financial impact of these common disorders has been substantial, with an estimated burden on direct and indirect health care costs of ~\$6 billion per year in the United States, with \$3 billion spent on hospitalizations, \$2 billion on physician office visits, and \$1 billion in decreased productivity and days lost from work.

■ GASTRIC PHYSIOLOGY

Gastric Anatomy The gastric epithelial lining consists of rugae that contain microscopic gastric pits, each branching into four or five gastric glands made up of highly specialized epithelial cells. The makeup of gastric glands varies with their anatomic location. Glands within the gastric cardia comprise <5% of the gastric gland area and contain mucous and endocrine cells. The 75% of gastric glands are found within the oxyntic mucosa and contain mucous neck, parietal, chief, endocrine, enterochromaffin, and enterochromaffin-like (ECL) cells (Fig. 324-1). Highly specialized tuft cells are located in the neck region of the gastric gland. These specialized cells are thought to sample luminal contents, which in turn may be important in regulating gastric acid secretion. Pyloric glands contain mucous and endocrine cells (including gastrin cells) and are found in the antrum.

The parietal cell, also known as the oxytic cell, is usually found in the neck or isthmus or in the oxyntic gland. The resting, or unstimulated, parietal cell has prominent cytoplasmic tubulovesicles and intracellular canalicular containing short microvilli along its apical surface (Fig. 324-2). H⁺,K⁺-adenosine triphosphatase (ATPase) is expressed in the tubulovesicle membrane; upon cell stimulation, this membrane, along with apical membranes, transforms into a dense network of apical intracellular canalicular containing long microvilli. Acid secretion, a process requiring high energy, occurs at the apical canalicular surface. Numerous mitochondria (30–40% of total cell volume) generate the energy required for secretion.

Gastroduodenal Mucosal Defense The gastric epithelium is under constant assault by a series of endogenous noxious factors, including hydrochloric acid (HCl), pepsinogen/pepsin, and bile salts. In addition, a steady flow of exogenous substances such as medications, alcohol, and bacteria encounter the gastric mucosa. A highly intricate biologic system is in place to provide defense from mucosal injury and to repair any injury that may occur.

The mucosal defense system can be envisioned as a three-level barrier, composed of preepithelial, epithelial, and subepithelial elements

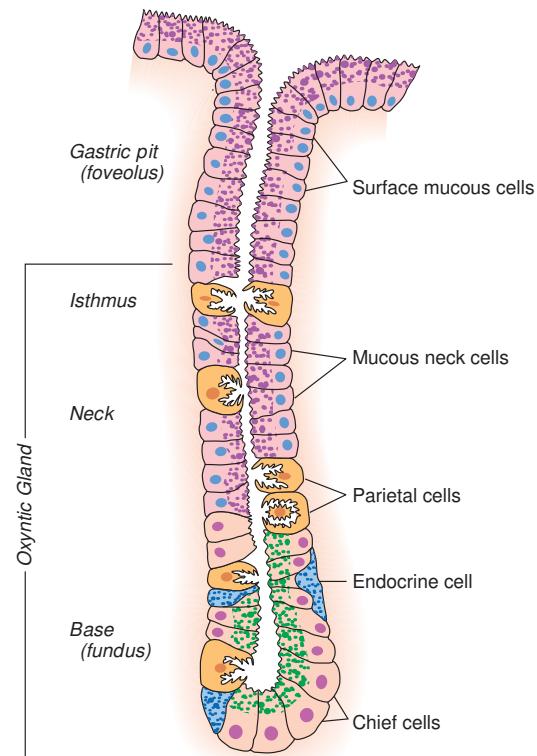


FIGURE 324-1 Diagrammatic representation of the oxyntic gastric gland. (Reproduced with permission from S Ito, RJ Winchester: The Fine Structure of the Gastric Mucosa in the Rat. *J Cell Biol* 16:541, 1963.)

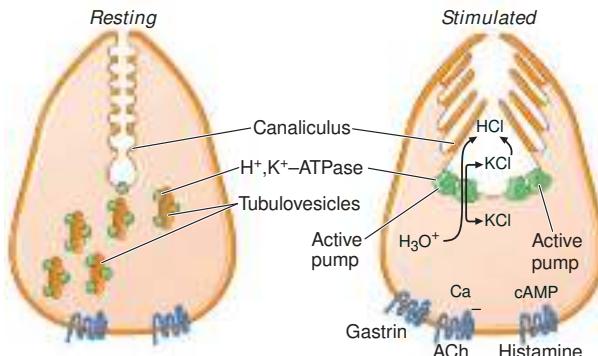


FIGURE 324-2 Gastric parietal cell undergoing transformation after secretagogue-mediated stimulation. cAMP, cyclic adenosine monophosphate. (Reproduced with permission from SJ Hersey, GSachs: *Gastric acid secretion*. Am Physiol Soc 75:155, 1995.)

(Fig. 324-3). The first line of defense is a mucus-bicarbonate-phospholipid layer, which serves as a physicochemical barrier to multiple molecules, including hydrogen ions. Mucus is secreted in a regulated fashion by gastroduodenal surface epithelial cells. It consists primarily of water (95%) and a mixture of phospholipids and glycoproteins (mucin). The mucous gel functions as a nonstirred water layer impeding diffusion of ions and molecules such as pepsin. Bicarbonate, secreted in a regulated manner by surface epithelial cells of the gastroduodenal mucosa into the mucous gel, forms a pH gradient ranging from 1 to 2 at the gastric luminal surface and reaching 6–7 along the epithelial cell surface.

Surface epithelial cells provide the next line of defense through several factors, including mucus production, epithelial cell ionic transporters that maintain intracellular pH and bicarbonate production, and intracellular tight junctions. Surface epithelial cells generate heat shock proteins that prevent protein denaturation and protect cells from certain factors such as increased temperature, cytotoxic agents, or oxidative stress. Epithelial cells also generate trefoil factor family peptides and cathelicidins, which also play a role in surface cell protection and regeneration. If the preepithelial barrier is breached, gastric epithelial cells bordering a site of injury can migrate to restore a damaged region (*restoration*). This process occurs independent of cell division and requires uninterrupted blood flow and an alkaline pH in the surrounding environment. Several growth factors, including epidermal growth factor (EGF), transforming growth factor (TGF) α , and basic fibroblast growth factor (FGF), modulate the process of restitution. Larger defects that are not effectively repaired by restitution require cell proliferation. Epithelial cell regeneration is regulated by prostaglandins and growth factors such as EGF and TGF- α . In tandem with epithelial cell renewal, formation of new vessels (*angiogenesis*) within the injured microvascular bed occurs. Both FGF and vascular endothelial growth factor (VEGF) are important in regulating angiogenesis in the gastric mucosa. In addition, the gastric peptide gastrin (see below) has been recently found to stimulate cell proliferation, migration, invasion, angiogenesis, and autophagy. Finally, gastric parietal cells (see below) express sonic hedgehog, a family of proteins important in regulating cell lineage in multiple organs. This latter finding suggests that parietal cells may also have the ability to regulate gastric stem cells.

An elaborate microvascular system within the gastric submucosal layer is the key component of the subepithelial defense/repair system, providing HCO_3^- , which neutralizes the acid generated by the parietal cell. Moreover, this microcirculatory bed provides an adequate supply of micronutrients and oxygen while removing toxic metabolic by-products. Several locally produced factors including nitric oxide (NO) (see below), hydrogen sulfide, and prostacyclin contribute to the vascular protective pathway through vasodilation of the microcirculation.

Prostaglandins play a central role in gastric epithelial defense/repair (Fig. 324-4). The gastric mucosa contains abundant levels of prostaglandins that regulate the release of mucosal bicarbonate and mucus, inhibit parietal cell secretion, and are important in maintaining

mucosal blood flow and epithelial cell restitution. Prostaglandins are derived from esterified arachidonic acid, which is formed from phospholipids (cell membrane) by the action of phospholipase A₂. A key enzyme that controls the rate-limiting step in prostaglandin synthesis is cyclooxygenase (COX), which is present in two isoforms (COX-1, COX-2), each having distinct characteristics regarding structure, tissue distribution, and expression. COX-1 is expressed in a host of tissues, including the stomach, platelets, kidneys, and endothelial cells. This isoform is expressed in a constitutive manner and plays an important role in maintaining the integrity of renal function, platelet aggregation, and gastrointestinal (GI) mucosal integrity. In contrast, the expression of COX-2 is inducible by inflammatory stimuli, and it is expressed in macrophages, leukocytes, fibroblasts, and synovial cells. The beneficial effects of nonsteroidal anti-inflammatory drugs (NSAIDs) on tissue inflammation are due to inhibition of COX-2; the toxicity of these drugs (e.g., GI mucosal ulceration and renal dysfunction) is related to inhibition of the COX-1 isoform. The highly COX-2-selective NSAIDs have the potential to provide the beneficial effect of decreasing tissue inflammation while minimizing toxicity in the GI tract. Selective COX-2 inhibitors have had adverse effects on the cardiovascular (CV) system, leading to increased risk of myocardial infarction. Therefore, the U.S. Food and Drug Administration (FDA) has removed two of these agents (valdecoxib and rofecoxib) from the market (see below).

NO is important in the maintenance of gastric mucosal integrity. The key enzyme NO synthase is constitutively expressed in the mucosa and contributes to cytoprotection by stimulating gastric mucus, increasing mucosal blood flow, and maintaining epithelial cell barrier function. The central nervous system (CNS) and hormonal factors also play a role in regulating mucosal defense through multiple pathways (Fig. 324-3).

Since the discovery of *Helicobacter pylori* and its impact on gastric pathology, it has become clear that the stomach has an elaborate and complex inherent immunologic system in place. Although a detailed description of the gastric immune system is beyond the scope of this chapter, several features are worth highlighting. The gastric immune response to certain pathogens such as *H. pylori* (see below) involves extensive interplay between innate (dendritic cells, epithelial cells, neutrophils, and macrophages) and adaptive (B and T cells) components. Helper T cells (T_H and T_H regulatory cells) have been extensively studied and appear to play an important role in a broad array of gastric physiology extending from gastric secretion to epithelial cell turnover via production of a number of cytokines.

The discovery of *H. pylori* has also led to the understanding that the stomach, once thought to be devoid of microorganisms due to its highly adverse environment (acid and pepsin), can serve as host for bacterial communities consisting of hundreds of phylotypes, otherwise known as its microbiota. The conceptual framework of the microbiome has been receiving extensive attention in light of its importance in human health and disease. The overall relevance of the gastric microbiome and its impact on gastric pathology remain to be established, but it is likely that alteration of microorganism homeostasis will play a role in aspects of certain disorders such as PUD, gastritis, and gastric cancer.

Physiology of Gastric Secretion HCl and pepsinogen are the two principal gastric secretory products capable of inducing mucosal injury. Gastric acid and pepsinogen play a physiologic role in protein digestion; absorption of iron, calcium, magnesium, and vitamin B₁₂; and killing ingested bacteria. Acid secretion should be viewed as occurring under basal and stimulated conditions. Basal acid production occurs in a circadian pattern, with the highest levels occurring during the night and lowest levels during the morning hours. Cholinergic input via the vagus nerve and histaminergic input from local gastric sources are the principal contributors to basal acid secretion. Stimulated gastric acid secretion occurs primarily in three phases based on the site where the signal originates (cephalic, gastric, and intestinal). Sight, smell, and taste of food are the components of the cephalic phase, which stimulates gastric secretion via the vagus nerve. The gastric phase is activated once food enters the stomach. This component of

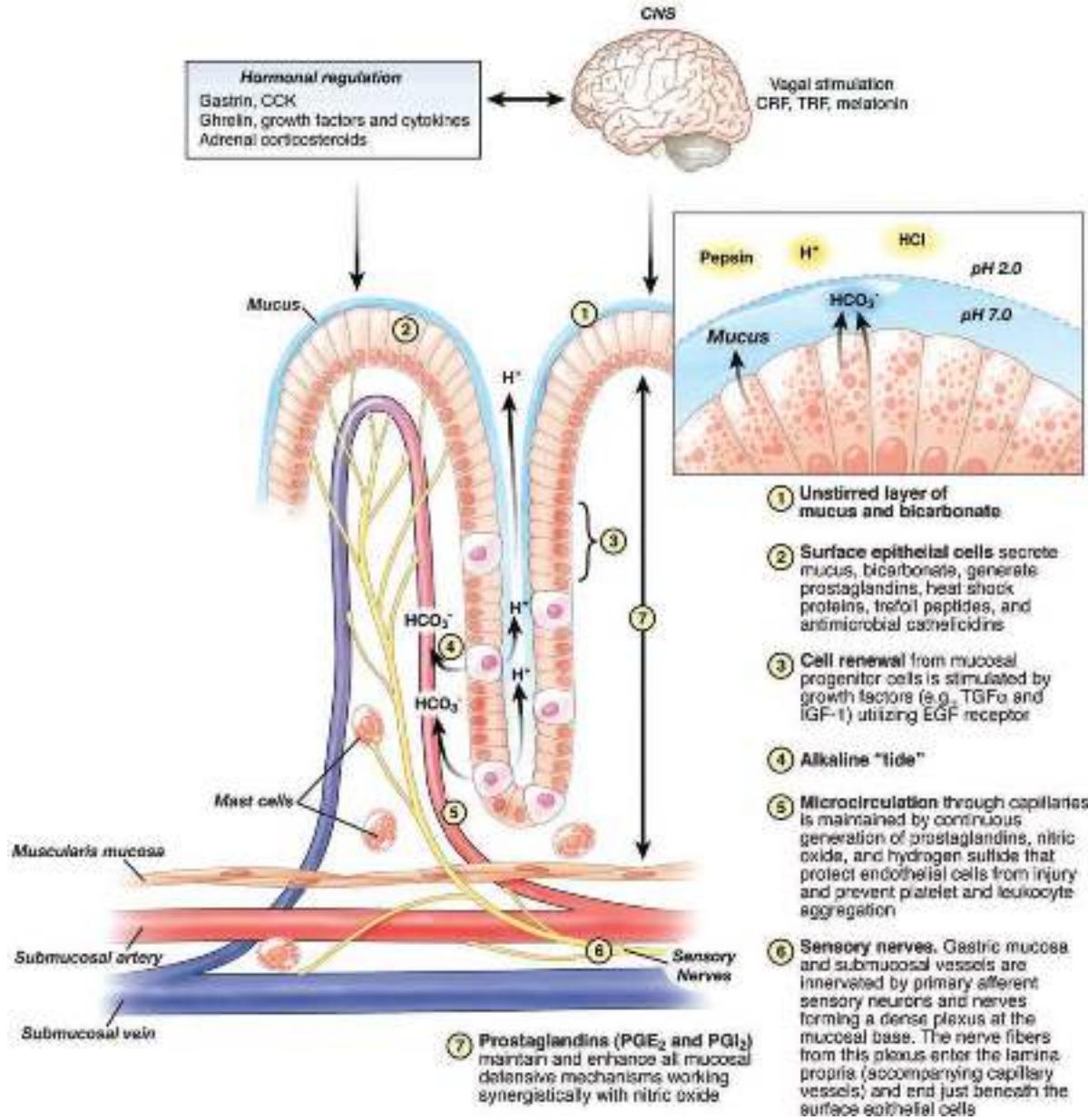


FIGURE 324-3 Components involved in providing gastroduodenal mucosal defense and repair. CCK, cholecystokinin; CRF, corticotropin-releasing factor; EGF, epidermal growth factor; HCl, hydrochloride; IGF, insulin-like growth factor; TGF α , transforming growth factor α ; TRF, thyrotropin releasing factor. (Republished with permission of John Wiley and Sons Inc, from Bioregulation and Its Disorders in the Gastrointestinal Tract, T Yoshikawa, T Arakawa [eds]: 1998; permission conveyed through Copyright Clearance Center, Inc.)

secretion is driven by nutrients (amino acids and amines) that directly (via peptone and amino acid receptors) and indirectly (via stimulation of intramural gastrin-releasing peptide neurons) stimulate the G cell to release gastrin, which in turn activates the parietal cell via direct and indirect mechanisms. Distention of the stomach wall also leads to gastrin release and acid production. The last phase of gastric acid secretion is initiated as food enters the intestine and is mediated by luminal distention and nutrient assimilation. A series of pathways that inhibit gastric acid production are also set into motion during these phases. The GI hormone somatostatin is released from endocrine cells found in the gastric mucosa (D cells) in response to HCl. Somatostatin can inhibit acid production by both direct (parietal cell) and indirect mechanisms (decreased histamine release from ECL cells, ghrelin release from Gr cells and gastrin release from G cells). Additional neural (central and

peripheral) and humoral (amylin, atrial natriuretic peptide [ANP], cholecystokinin, ghrelin, interleukin 11 [IL-11], obestatin, secretin, and serotonin) factors play a role in counterbalancing acid secretion. Under physiologic circumstances, these phases occur simultaneously. Ghrelin, the appetite-regulating hormone expressed in Gr cells in the stomach, and its related peptide motilin (released from the duodenum) may increase gastric acid secretion through stimulation of histamine release from ECL cells, but this remains to be confirmed.

The acid-secreting parietal cell is located in the oxytic gland, adjacent to other cellular elements (ECL cell, D cell) important in the gastric secretory process (Fig. 324-5). This unique cell also secretes intrinsic factor (IF) and IL-11. The parietal cell expresses receptors for several stimulants of acid secretion, including histamine (H_2), gastrin (cholecystokinin 2/gastrin receptor), and acetylcholine (muscarinic,

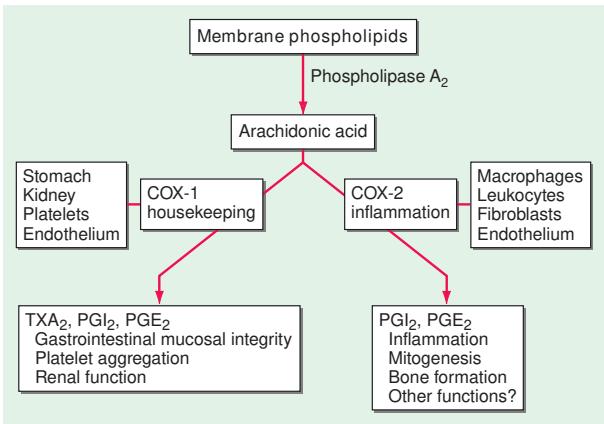


FIGURE 324-4 Schematic representation of the steps involved in synthesis of prostaglandin E₂ (PGE₂) and prostacyclin (PGI₂). Characteristics and distribution of the cyclooxygenase (COX) enzymes 1 and 2 are also shown. TXA₂, thromboxane A₂.

M₃). Binding of histamine to the H₂ receptor leads to activation of adenylate cyclase and the phosphoinositol pathways, in turn resulting in an increase in cyclic adenosine monophosphate (AMP) and intracellular calcium, respectively. Activation of the gastrin and muscarinic receptors results in activation of the protein kinase C/phosphoinositide signaling pathway. Each of these signaling pathways in turn regulates a series of downstream kinase cascades that control the acid-secreting pump, H⁺K⁺-ATPase. The discovery that different ligands and their corresponding receptors lead to activation of different signaling pathways explains the potentiation of acid secretion that occurs when histamine and gastrin or acetylcholine are combined. More importantly, this observation explains why blocking one receptor type (H₂) decreases acid secretion stimulated by agents that activate a different pathway (gastrin, acetylcholine). Parietal cells also express receptors for ligands that inhibit acid production (glucagon-like peptide-1, prostaglandins, somatostatin, EGF, neurotensin, and urocortin). Histamine also stimulates gastric acid secretion indirectly by activating the histamine H₃ receptor on D cells, which inhibits somatostatin release.

The enzyme H⁺K⁺-ATPase is responsible for generating the large concentration of H⁺. It is a membrane-bound protein that consists of two subunits, α and β . The active catalytic site is found within the α subunit; the function of the β subunit is unclear. This enzyme uses the chemical energy of adenosine triphosphate (ATP) to transfer H⁺ ions

from parietal cell cytoplasm to the secretory canalliculi in exchange for K⁺. The H⁺,K⁺-ATPase is located within the secretory canalliculi and in nonsecretory cytoplasmic tubulovesicles. The tubulovesicles are impermeable to K⁺, which leads to an inactive pump in this location. The distribution of pumps between the nonsecretory vesicles and the secretory canalliculus varies according to parietal cell activity (Fig. 324-2). Proton pumps are recycled back to the inactive state in cytoplasmic vesicles once parietal cell activation ceases. Ezrin (an actin binding protein), actin, myosin, soluble N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs), small G proteins of the Rab family, and secretory carrier membrane proteins (SCAMPS) are postulated to participate in parietal cell membrane translocation. In addition, acid secretion requires a number of apical and basolateral parietal cell membrane chloride and potassium channels. Parietal cells also express members of the sonic hedgehog (Shh) family proteins, which play an important role in regulating cell types in multiple organs. This family of proteins may also regulate cell differentiation as well as restitution of mucosal defense in the gastric epithelium.

The chief cell, found primarily in the gastric fundus, synthesizes and secretes pepsinogen, the inactive precursor of the proteolytic enzyme pepsin. The acid environment within the stomach leads to cleavage of the inactive precursor to pepsin and provides the low pH (<2) required for pepsin activity. Pepsin activity is significantly diminished at a pH of 4 and irreversibly inactivated and denatured at a pH of ≥7. Many of the secretagogues that stimulate acid secretion also stimulate pepsinogen release. The precise role of pepsin in the pathogenesis of PUD remains to be established.

■ PATHOPHYSIOLOGIC BASIS OF PUD

PUD encompasses both gastric ulcers (GUs) and duodenal ulcers (DUs). *Ulcers* are defined as breaks in the mucosal surface >5 mm in size, with depth to the submucosa. DUs and GUs share many common features in terms of pathogenesis, diagnosis, and treatment, but several factors distinguish them from one another. *H. pylori* and NSAIDs are the most common risk factors for PUD, with estimated odds ratios in the United States of 3.7 and 3.3, respectively. Additional risk factors (odds ratio) include chronic obstructive lung disease (2.34), chronic renal insufficiency (2.29), current tobacco use (1.99), former tobacco use (1.55), older age (1.67), three or more doctor visits in a year (1.49), coronary heart disease (1.46), former alcohol use (1.29), African-American race (1.20), obesity (1.18), and diabetes (1.13). Selective serotonin reuptake inhibitors (SSRIs) and gastric bypass surgery are also associated with an increased incidence of PUD. A rise in idiopathic PUD has also been noted. The mechanisms by which some of these risk factors lead to ulcer disease are highlighted below.

Epidemiology • DUODENAL ULCERS

DUs are estimated to occur in 6–15% of the Western population. The incidence of DUs declined steadily from 1960 to 1980 and has remained stable since then. The death rates, need for surgery, and physician visits have decreased by >50% over the past 30 years. The reason for the reduction in the frequency of DUs is likely related to the decreasing frequency of *H. pylori* in turn associated with overall improved sanitary conditions across the world. Before the discovery of *H. pylori*, the natural history of DUs was typified by frequent recurrences after initial therapy. Eradication of *H. pylori* has reduced these recurrence rates by >80%.

GASTRIC ULCERS GUs tend to occur later in life than duodenal lesions, with a peak incidence reported in the sixth decade. More than one-half of GUs occur

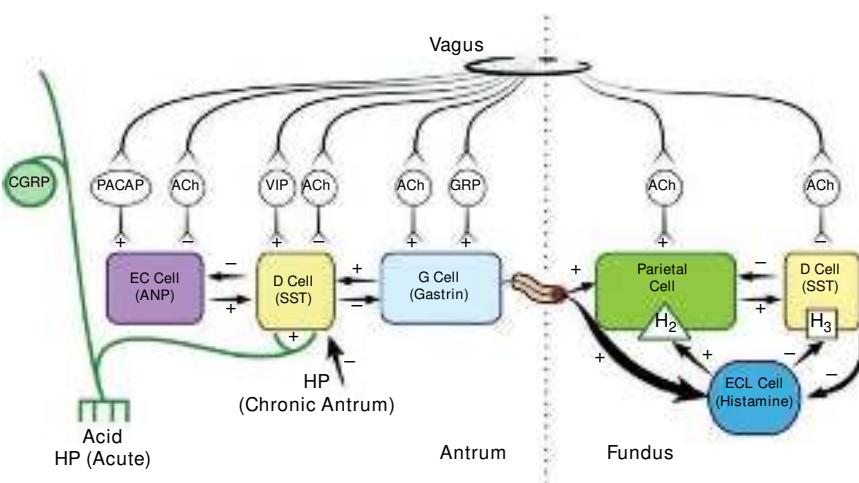


FIGURE 324-5 Regulation of gastric acid secretion at the cellular level. ACh, acetylcholine; ANP, atrial natriuretic peptide; CGRP, calcitonin gene-related peptide; EC, enterochromaffin; ECL, enterochromaffin-like; GRP, gastrin-releasing peptide; PACAP, pituitary adenylate-cyclase activating peptide; SST, somatostatin; VIP, vasoactive intestinal peptide.

in males and are less common than DUs, perhaps due to the higher likelihood of GUs being silent and presenting only after a complication develops. Autopsy studies suggest a similar incidence of DUs and GUs.

Pathology • DUODENAL ULCERS DUs occur most often in the first portion of the duodenum (>95%), with ~90% located within 3 cm of the pylorus. They are usually ≤ 1 cm in diameter but can occasionally reach 3–6 cm (giant ulcer). Ulcers are sharply demarcated, with depth at times reaching the muscularis propria. The base of the ulcer often consists of a zone of eosinophilic necrosis with surrounding fibrosis. Malignant DUs are extremely rare.

GASTRIC ULCERS In contrast to DUs, GUs can represent a malignancy and should be biopsied upon discovery. Benign GUs are most often found distal to the junction between the antrum and the acid secretory mucosa. Benign GUs are quite rare in the gastric fundus and are histologically similar to DUs. Benign GUs associated with *H. pylori* are also associated with antral gastritis. In contrast, NSAID-related GUs are not accompanied by chronic active gastritis but may instead have evidence of a chemical gastropathy, typified by foveolar hyperplasia, edema of the lamina propria, and epithelial regeneration in the absence of *H. pylori*. Extension of smooth-muscle fibers into the upper portions of the mucosa, where they are not typically found, may also occur.

Pathophysiology • DUODENAL ULCERS *H. pylori* and NSAID-induced injuries account for the majority of DUs. Many acid secretory abnormalities have been described in DU patients. Of these, average basal and nocturnal gastric acid secretion appears to be increased in DU patients as compared to controls; however, the level of overlap between DU patients and control subjects is substantial. The reason for this altered secretory process is unclear, but *H. pylori* infection may contribute. Bicarbonate secretion is significantly decreased in the duodenal bulb of patients with an active DU as compared to control subjects. *H. pylori* infection may also play a role in this process (see below).

GASTRIC ULCERS As in DUs, the majority of GUs can be attributed to either *H. pylori* or NSAID-induced mucosal damage. Prepyloric GUs or those in the body associated with a DU or a duodenal scar are similar in pathogenesis to DUs. Gastric acid output (basal and stimulated) tends to be normal or decreased in GU patients. When GUs develop in the presence of minimal acid levels, impairment of mucosal defense factors may be present. GUs have been classified based on their location: type I occur in the gastric body and tend to be associated with low gastric acid production; type II occur in the antrum, and gastric acid can vary from low to normal; type III occur within 3 cm of the pylorus and are commonly accompanied by DUs and normal or high gastric acid production; and type IV are found in the cardia and are associated with low gastric acid production.

***H. PYLORI* AND ACID PEPTIC DISORDERS** Gastric infection with the bacterium *H. pylori* accounts for the majority of PUD (Chap. 163). This organism also plays a role in the development of gastric mucosa-associated lymphoid tissue (MALT) lymphoma and gastric adenocarcinoma. Although the entire genome of *H. pylori* has been sequenced, it is still not clear how this organism, which resides in the stomach, causes ulceration in the duodenum. *H. pylori* eradication efforts may lead to a decrease in gastric cancer in high-risk populations, particularly in individuals who have not developed chronic atrophic gastritis and gastric metaplasia.

The Bacterium The bacterium, initially named *Campylobacter pyloridis*, is a gram-negative microaerophilic rod found most commonly in the deeper portions of the mucous gel coating the gastric mucosa or between the mucous layer and the gastric epithelium. It may attach to gastric epithelium but under normal circumstances does not appear to invade cells. It is strategically designed to live within the aggressive environment of the stomach. It is S-shaped (~0.5–3 μm in size) and contains multiple sheathed flagella. Initially, *H. pylori* resides in the antrum but, over time, migrates toward the more proximal segments of the stomach. The organism is capable of transforming into a coccoid form, which represents a dormant state that may facilitate survival in adverse conditions. The genome of *H. pylori* (1.65 million base pairs)

encodes ~1500 proteins. Among this multitude of proteins there are factors that are essential determinants of *H. pylori*-mediated pathogenesis and colonization such as the outer membrane protein (Hop proteins), urease, and the vacuolating cytotoxin (Vac A). Moreover, the majority of *H. pylori* strains contain a genomic fragment that encodes the cag pathogenicity island (cag-PAI). Several of the genes that make up cag-PAI encode components of a type IV secretion island that translocates Cag A into host cells. Once in the cell, Cag A activates a series of cellular events important in cell growth and cytokine production. *H. pylori* also has extensive genetic diversity that in turn enhances its ability to promote disease. The first step in infection by *H. pylori* is dependent on the bacteria's motility and its ability to produce urease. Urease produces ammonia from urea, an essential step in alkalinizing the surrounding pH. Additional bacterial factors include catalase, lipase, adhesins, platelet-activating factor, and pic B (induces cytokines). Multiple strains of *H. pylori* exist and are characterized by their ability to express several of these factors (Cag A, Vac A, etc.). It is possible that the different diseases related to *H. pylori* infection can be attributed to different strains of the organism with distinct pathogenic features.

Epidemiology The prevalence of *H. pylori* varies throughout the world and depends largely on the overall standard of living in the region. In developing parts of the world, 80% of the population may be infected by the age of 20, whereas the prevalence is 20–50% in industrialized countries. In contrast, in the United States, this organism is rare in childhood. The overall prevalence of *H. pylori* in the United States is ~30%, with individuals born before 1950 having a higher rate of infection than those born later. About 10% of Americans <30 years of age are colonized with the bacteria. The rate of infection with *H. pylori* in industrialized countries has decreased substantially in recent decades. The steady increase in the prevalence of *H. pylori* noted with increasing age is due primarily to a cohort effect, reflecting higher transmission during a period in which the earlier cohorts were children. It has been calculated through mathematical models that improved sanitation during the latter half of the nineteenth century dramatically decreased transmission of *H. pylori*. Moreover, with the present rate of intervention, the organism will be ultimately eliminated from the United States. Two factors that predispose to higher colonization rates include poor socioeconomic status and less education. These factors, not race, are responsible for the rate of *H. pylori* infection in blacks and Hispanic Americans being double the rate seen in whites of comparable age. Other risk factors for *H. pylori* infection are (1) birth or residence in a developing country, (2) domestic crowding, (3) unsanitary living conditions, (4) unclean food or water, and (5) exposure to gastric contents of an infected individual.

Transmission of *H. pylori* occurs from person to person, following an oral-oral or fecal-oral route. The risk of *H. pylori* infection is declining in developing countries. The rate of infection in the United States has fallen by >50% when compared to 30 years ago.

Pathophysiology *H. pylori* infection is virtually always associated with a chronic active gastritis, but only 10–15% of infected individuals develop frank peptic ulceration. The basis for this difference is unknown but is likely due to a combination of host and bacterial factors, some of which are outlined below. Initial studies suggested that >90% of all DUs were associated with *H. pylori*, but *H. pylori* is present in only 30–60% of individuals with GUs and 50–70% of patients with DUs. The pathophysiology of ulcers not associated with *H. pylori* or NSAID ingestion (or the rare Zollinger-Ellison syndrome [ZES]) is becoming more relevant as the incidence of *H. pylori* is dropping, particularly in the Western world (see below).

The particular end result of *H. pylori* infection (gastritis, PUD, gastric MALT lymphoma, gastric cancer) is determined by a complex interplay between bacterial and host factors (Fig. 324-6).

Bacterial factors: *H. pylori* is able to facilitate gastric residence, induce mucosal injury, and avoid host defense. Different strains of *H. pylori* produce different virulence factors including γ-glutamyl transpeptidase (GGT), cytotoxin-associated gene A (Cag A) product, and virulence components vacuolating toxin (Vac A), in addition to

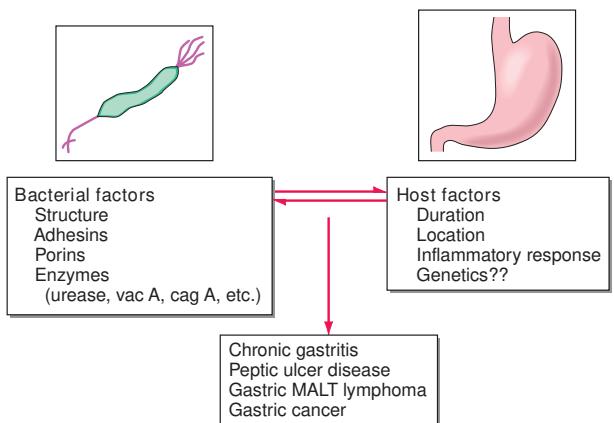


FIGURE 324-6 Outline of the bacterial and host factors important in determining *H. pylori*-induced gastrointestinal disease. MALT, mucosal-associated lymphoid tissue.

pathogen-associated molecular patterns (PAMPs) such as flagella and lipopolysaccharide (LPS). A specific region of the bacterial genome, the pathogenicity island (cag-PAI), encodes the virulence factors Cag A and pic B. Vac A also contributes to pathogenicity, although it is not encoded within the pathogenicity island. These virulence factors, in conjunction with additional bacterial constituents, can cause mucosal damage, in part through their ability to target the host immune cells. For example, Vac A targets human CD4 T cells, inhibiting their proliferation, and in addition can disrupt normal function of B cells, CD8 T cells, macrophages, and mast cells. Multiple studies have demonstrated that *H. pylori* strains that are cag-PAI positive are associated with a higher risk of PUD, premalignant gastric lesions, and gastric cancer than are strains that lack the cag-PAI. In addition, *H. pylori* may directly inhibit parietal cell H^+,K^+ -ATPase activity through a Cag A-dependent mechanism, leading in part to the low acid production observed after acute infection with the organism. Urease, which allows the bacteria to reside in the acidic stomach, generates NH_3 , which can damage epithelial cells. The bacteria produce surface factors that are chemotactic for neutrophils and monocytes, which in turn contribute to epithelial cell injury (see below). *H. pylori* makes proteases and phospholipases that break down the glycoprotein lipid complex of the mucous gel, thus reducing the efficacy of this first line of mucosal defense. *H. pylori* expresses adhesins (outer membrane proteins like BabA), which facilitate attachment of the bacteria to gastric epithelial cells. Although LPS of gram-negative bacteria often plays an important role in the infection, *H. pylori* LPS has low immunologic activity compared to that of other organisms. It may promote a smoldering chronic inflammation.

Host factors: Studies in twins suggest that there may be genetic predisposition to acquire *H. pylori*. The inflammatory response to *H. pylori* includes recruitment of neutrophils, lymphocytes (T and B), macrophages, and plasma cells. The pathogen leads to local injury by binding to class II major histocompatibility complex (MHC) molecules expressed on gastric epithelial cells, leading to cell death (*apoptosis*). Moreover, bacterial strains that encode cag-PAI can introduce Cag A into the host cells, leading to further cell injury and activation of cellular pathways involved in cytokine production and repression of tumor-suppressor genes. Elevated concentrations of multiple cytokines are found in the gastric epithelium of *H. pylori*-infected individuals, including interleukin (IL) 1 α / β , IL-2, IL-6, IL-8, tumor necrosis factor (TNF) α , and interferon (IFN) γ . *H. pylori* infection also leads to both a mucosal and a systemic humoral response, which does not lead to eradication of the bacteria but further compounds epithelial cell injury. Additional mechanisms by which *H. pylori* may cause epithelial cell injury include (1) activated neutrophil-mediated production of reactive oxygen or nitrogen species and enhanced epithelial cell turnover and (2) apoptosis related to interaction with T cells (T helper 1 [T_{H}^1] cells) and IFN- γ . Finally, the human stomach is colonized by a host

of commensal organisms that may affect the likelihood of *H. pylori* infection and subsequent mucosal injury. Moreover, colonization of the stomach with *H. pylori* likely alters the composition of the gastric microbiota. The impact of the latter on gastric pathophysiology remains unknown. *H. pylori* also appears to regulate NO formation via different mechanisms that in turn may contribute to the organism's cytotoxic effects. Specifically, *H. pylori*-derived factors, such as urease, or the bacterium itself, stimulate NO synthase (NOS2) expression in macrophages and in gastric epithelial cells leading to NO release and subsequent cytotoxic effect on surrounding cells. *H. pylori* also leads to the formation of 8-nitroguanine (8-NO₂-Gua), which in conjunction with oncoprotein Cag A, may contribute to the development of gastric cancer.

The reason for *H. pylori*-mediated duodenal ulceration remains unclear. Studies suggest that *H. pylori* associated with duodenal ulceration may be more virulent. In addition, certain specific bacterial factors such as the DU-promoting gene A (*dupA*) may be associated with the development of DUs. Another potential contributing factor is that gastric metaplasia in the duodenum of DU patients, which may be due to high acid exposure (see below), permits *H. pylori* to bind to it and produce local injury secondary to the host response. Another hypothesis is that *H. pylori* antral infection could lead to increased acid production, increased duodenal acid, and mucosal injury. Basal and stimulated (meal, gastrin-releasing peptide [GRP]) gastrin release is increased in *H. pylori*-infected individuals, and somatostatin-secreting D cells may be decreased. *H. pylori* infection might induce increased acid secretion through both direct and indirect actions of *H. pylori* and proinflammatory cytokines (IL-8, TNF, and IL-1) on G, D, and parietal cells (Fig. 324-7). GUs, in contrast, are associated with *H. pylori*-induced pangastritis and normal or low gastric acid secretion. The *H. pylori*-mediated decrease in gastric acid secretion after long-term infection may be due to the bacterium's ability to inhibit H^+,K^+ -ATPase expression. *H. pylori* infection has also been associated with decreased duodenal mucosal bicarbonate production. Data supporting and contradicting each of these interesting theories have been demonstrated. Thus, the mechanism by which *H. pylori* infection of the stomach leads to duodenal ulceration remains to be established. The development of in vitro organoids, a unique tool that replicates in part the multicellular structure of the intact organ, provides a more physiologic model for experimentation in an in vitro system. Moreover, the development of advanced microscopic optical imaging techniques

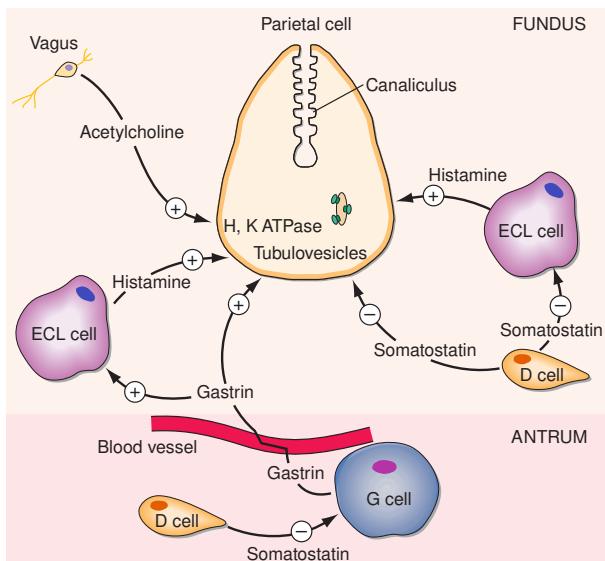


FIGURE 324-7 Summary of potential mechanisms by which *H. pylori* may lead to gastric secretory abnormalities. D, somatostatin cell; ECL, enterochromaffin-like cell; G, G cell. (Reproduced with permission from J Calam et al: How does *Helicobacter pylori* cause mucosal damage? Its effect on acid and gastrin physiology. *Gastroenterology* 113:543, 1997.)

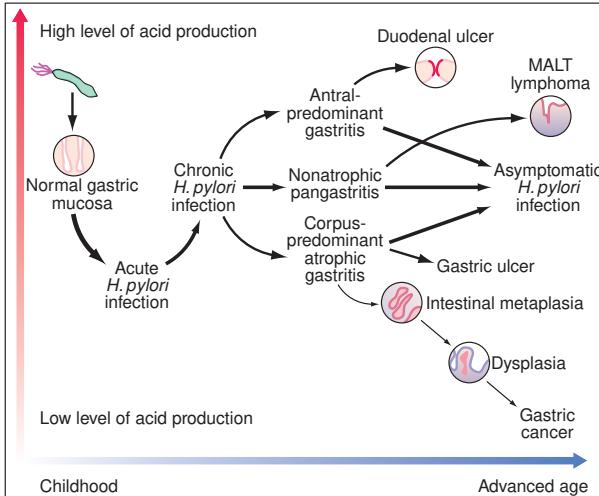


FIGURE 324-8 Natural history of *H. pylori* infection. MALT, mucosal-associated lymphoid tissue. (From S Suerbaum, P Michetti: *Helicobacter pylori* infection. *N Engl J Med* 347:1175, 2002. Copyright © 2002 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

will lead to increased understanding of parietal cell adaptation to *H. pylori* infection.

In summary, the final effect of *H. pylori* on the GI tract is variable and determined by microbial and host factors. The type and distribution of gastritis correlate with the ultimate gastric and duodenal pathology observed. Specifically, the presence of antral-predominant gastritis is associated with DU formation; gastritis involving primarily the corpus predisposes to the development of GUs, gastric atrophy, and ultimately gastric carcinoma (Fig. 324-8).

NSAID-INDUCED DISEASE

Epidemiology NSAIDs represent a group of the most commonly used medications in the world and the United States. It is estimated that 7 billion dollars per year are spent on NSAIDs worldwide, with >30 billion over-the-counter tablets sold. More than 30 million individuals take NSAIDs, with >100 million prescriptions sold yearly in the United States alone. In fact, after the introduction of COX-2 inhibitors in the year 2000, the number of prescriptions written for NSAIDs was >111 million at a cost of \$4.8 billion. Side effects and complications due to NSAIDs are considered the most common drug-related toxicities in the United States. The spectrum of NSAID-induced morbidity ranges from nausea and dyspepsia (prevalence reported as high as 50–60%) to a serious GI complication such as endoscopy-documented peptic ulceration (15–30% of individuals taking NSAIDs regularly), which is complicated by bleeding or perforation in as many as 1.5% of users per year. It is estimated that NSAID-induced GI bleeding accounts for 60,000–120,000 hospital admissions per year, and deaths related to NSAID-induced toxicity may be as high as 16,000 per year in the United States. Approximately 4–5% of patients develop symptomatic ulcers within 1 year. Unfortunately, dyspeptic symptoms do not correlate with NSAID-induced pathology. Over 80% of patients with serious NSAID-related complications did not have preceding dyspepsia. In view of the lack of warning signs, it is important to identify patients who are at increased risk for morbidity and mortality related to NSAID usage. Even 75 mg/d of aspirin may lead to serious GI ulceration; thus, no dose of NSAID is completely safe. In fact, the incidence of mucosal injury (ulcers and erosions) in patients taking low-dose aspirin (75–325 mg) has been estimated to range from as low as 8 to as high as 60%. It appears that *H. pylori* infection increases the risk of PUD-associated GI bleeding in chronic users of low-dose aspirin. Established risk factors include advanced age, history of ulcer, concomitant use of glucocorticoids, high-dose NSAIDs, multiple NSAIDs, concomitant use of anticoagulants or clopidogrel, and serious or multisystem disease. Possible risk factors include concomitant infection with *H. pylori*, cigarette smoking, and

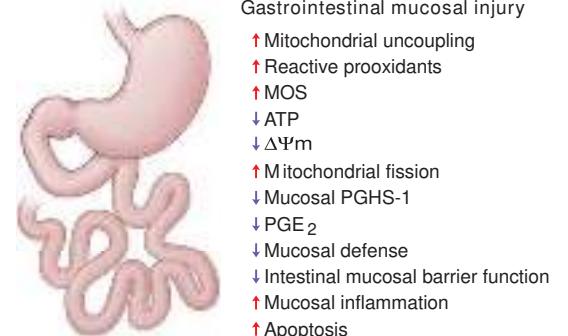


FIGURE 324-9 Effect of nonsteroidal anti-inflammatory drugs (NSAIDs) on different target organs. The action of NSAIDs on major organs including stomach, small intestine, heart, liver, kidney, respiratory tract, and brain is mainly mediated through prostaglandin endoperoxide synthase (PGHS)-dependent prostanoid modulation and alteration of mitochondrial functional integrity leading to mitochondrial oxidative stress (MOS) generation, depolarization of mitochondrial transmembrane potential ($\Delta\Psi_m$), and consequent cell death. However, in heart, low-dose aspirin actually offers cardioprotection through antithrombotic effect. Upward arrows indicate upregulation/elevation; downward arrows indicate downregulation/depletion. (From S Bindu et al: Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. *Biochem Pharmacol* 180:114147, 2020.)

alcohol consumption. SSRIs have a synergistic effect on the induction of GI bleeding believed to be due in part to this agent's ability to decrease platelet aggregation by decreasing serotonin content in platelets.

Pathophysiology Prostaglandins play a critical role in maintaining gastroduodenal mucosal integrity and repair. It therefore follows that interruption of prostaglandin synthesis can impair mucosal defense and repair, thus facilitating mucosal injury via a systemic mechanism. Animal studies have demonstrated that neutrophil adherence to the gastric microcirculation plays an essential role in the initiation of NSAID-induced mucosal injury. A summary of the pathogenetic pathways by which systemically administered NSAIDs may lead to mucosal injury is shown in Fig. 324-9. Single nucleotide polymorphisms (SNPs) have been found in several genes, including those encoding certain subtypes of cytochrome P450 (see below), IL-1 β (*IL-1 β*), angiotensinogen (*AGT*), and an organic ion transporting polypeptide (*SLCO1B1*), but these findings need confirmation in larger-scale studies.

Injury to the mucosa also occurs as a result of the topical use of NSAIDs, leading to increased epithelial surface permeability. Aspirin and many NSAIDs are weak acids that remain in a nonionized lipophilic form when found within the acid environment of the stomach. Under these conditions, NSAIDs migrate across lipid membranes of epithelial cells, leading to cell injury once trapped intracellularly in an ionized form. Topical NSAIDs can also alter the surface mucus layer, permitting back diffusion of H⁺ and pepsin, leading to further epithelial cell damage. Moreover, enteric-coated or buffered preparations are also associated with risk of peptic ulceration. NSAIDs can also lead to mucosal injury via production of additional proinflammatory mediators such as TNF and leukotrienes through simultaneous activation of the lipoxygenase pathway.

The interplay between *H. pylori* and NSAIDs in the pathogenesis of PUD is complex. Meta-analysis supports the conclusion that each of these aggressive factors is an independent and synergistic risk factor for PUD and its complications such as GI bleeding. For example, eradication of *H. pylori* reduces the likelihood of GI complications in high-risk individuals to levels observed in individuals with average risk of NSAID-induced complications.

In summary, NSAID-induced mucosal injury is a multifaceted process involving the interaction of multiple, often synergistic pathophysiologic processes at the epithelium and surrounding interfaces.

PATHOGENETIC FACTORS UNRELATED TO *H. PYLORI* AND NSAIDS IN ACID PEPTIC DISEASE Cigarette smoking has been implicated in the pathogenesis of PUD. Not only have smokers been found to have ulcers more frequently than do nonsmokers, but smoking appears to decrease healing rates, impair response to therapy, and increase

ulcer-related complications such as perforation. The mechanism responsible for increased ulcer diathesis in smokers is unknown. Theories have included altered gastric emptying, decreased proximal duodenal bicarbonate production, increased risk for *H. pylori* infection, and cigarette-induced generation of noxious mucosal free radicals. Genetic predisposition may play a role in ulcer development. First-degree relatives of DU patients are three times as likely to develop an ulcer; however, the potential role of *H. pylori* infection in contacts is a major consideration. Increased frequencies of blood group O and of the nonsecretor status have also been implicated as genetic risk factors for peptic diathesis. However, *H. pylori* preferentially binds to group O antigens. Additional genetic factors have been postulated to predispose certain individuals to developing PUD and/or upper GI bleeding. Specifically, genes encoding the NSAID-metabolizing enzymes cytochrome P450 2C9 and 2C8 (CYP2C9 and CYP2C8) are potential susceptibility genes for NSAID-induced PUD, but unfortunately, the studies have not been consistent in demonstrating this association. In a United Kingdom study, the *CYP2C19*17* gain-of-function polymorphism was associated with PUD in a Caucasian cohort, irrespective of ulcer etiology. These findings need to be confirmed in broader studies. Psychological stress has been thought to contribute to PUD, but studies examining the role of psychological factors in its pathogenesis have generated conflicting results. Although PUD is associated with certain personality traits (neuroticism), these same traits are also present in individuals with nonulcer dyspepsia (NUD) and other functional and organic disorders.

Diet has also been thought to play a role in peptic diseases. Certain foods and beverages can cause dyspepsia, but no convincing studies indicate an association between ulcer formation and a specific diet. Specific chronic disorders have been shown to have a strong association with PUD: (1) advanced age, (2) chronic pulmonary disease, (3) chronic renal failure, (4) cirrhosis, (5) nephrolithiasis, (6) α_1 -antitrypsin deficiency, and (7) systemic mastocytosis. Disorders with a possible association are (1) hyperparathyroidism, (2) coronary artery disease, (3) polycythemia vera, (4) chronic pancreatitis, (5) former alcohol use, (6) obesity, (7) African-American race, and (8) three or more doctor visits in a year.

Multiple factors play a role in the pathogenesis of PUD. The two predominant causes are *H. pylori* infection and NSAID ingestion. PUD not related to *H. pylori* or NSAIDs is increasing. Other less common causes of PUD are shown in Table 324-1. These etiologic agents should be considered as the incidence of *H. pylori* is decreasing. Independent of the inciting or injurious agent, peptic ulcers develop as a result of an imbalance between mucosal protection/repair and aggressive factors. Gastric acid plays an important role in mucosal injury.

■ CLINICAL FEATURES

History Abdominal pain is common to many GI disorders, including DU and GU, but it has a poor predictive value for the presence of either DU or GU. Approximately two-thirds of patients with PUD do not have abdominal pain, and up to 87% of patients with NSAID-induced mucosal disease can present with a complication (bleeding, perforation, and obstruction) without antecedent symptoms. Despite this poor correlation, a careful history and physical examination are essential components of the approach to a patient suspected of having peptic ulcers.

Epigastric pain described as a burning or gnawing discomfort can be present in both DU and GU. The discomfort is also described as an ill-defined, aching sensation or as hunger pain. The typical pain pattern in DU occurs 90 min to 3 h after a meal and is frequently relieved by antacids or food. Pain that awakes the patient from sleep (between midnight and 3 A.M.) is the most discriminating symptom, with two-thirds of DU patients describing this complaint. Unfortunately, this symptom is also present in one-third of patients with NUD (see below). Elderly patients are less likely to have abdominal pain as a manifestation of PUD and may instead present with a complication such as ulcer bleeding or perforation. The pain pattern in GU patients may be different from that in DU patients, where discomfort may actually be precipitated by food. Nausea and weight loss occur more commonly in

TABLE 324-1 Causes of Ulcers Not Caused by *Helicobacter pylori* and NSAIDs

Pathogenesis of Non-Hp and Non-NSAID Ulcer Disease

Infection

- Cytomegalovirus
- Herpes simplex virus
- Helicobacter heilmannii*

Drug/Toxin

- Bisphosphonates
- Chemotherapy
- Clopidogrel
- Crack cocaine
- Glucocorticoids (when combined with NSAIDs)
- Mycophenolate mofetil
- Potassium chloride

Miscellaneous

- Basophilia in myeloproliferative disease
- Duodenal obstruction (e.g., annular pancreas)
- Infiltrating disease
- Ischemia
- Radiation therapy
- Eosinophilic infiltration
- Sarcoidosis
- Crohn's disease
- Idiopathic hypersecretory state

Abbreviations: Hp, *H. pylori*; NSAIDs, nonsteroidal anti-inflammatory drugs.

GU patients. Endoscopy detects ulcers in <30% of patients who have dyspepsia.

The mechanism for development of abdominal pain in ulcer patients is unknown. Several possible explanations include acid-induced activation of chemical receptors in the duodenum, enhanced duodenal sensitivity to bile acids and pepsin, and altered gastroduodenal motility.

Variation in the intensity or distribution of the abdominal pain, as well as the onset of associated symptoms such as nausea and/or vomiting, may be indicative of an ulcer complication. Dyspepsia that becomes constant, is no longer relieved by food or antacids, or radiates to the back may indicate a penetrating ulcer (pancreas). Sudden onset of severe, generalized abdominal pain may indicate perforation. Pain worsening with meals, nausea, and vomiting of undigested food suggest gastric outlet obstruction. Tarry stools or coffee-ground emesis indicate bleeding.

Physical Examination Epigastric tenderness is the most frequent finding in patients with GU or DU. Pain may be found to the right of the midline in 20% of patients. Unfortunately, the predictive value of this finding is low. Physical examination is critically important for discovering evidence of ulcer complication. Tachycardia and orthostasis suggest dehydration secondary to vomiting or active GI blood loss. A severely tender, board-like abdomen suggests a perforation. Presence of a succussion splash indicates retained fluid in the stomach, suggesting gastric outlet obstruction.

PUD-Related Complications

GASTROINTESTINAL BLEEDING GI bleeding is the most common complication observed in PUD. Bleeding is estimated to occur in 19.4–57 per 100,000 individuals in a general population or in ~15% of patients. Bleeding and complications of ulcer disease occur more often in individuals >60 years of age. The 30-day mortality rate is as high as 2.5–10%. The higher incidence in the elderly is likely due to the increased use of NSAIDs in this group. In addition, up to 80% of the mortality in PUD-related bleeding is due to nonbleeding causes such as multiorgan failure (24%), pulmonary complications (24%), and malignancy (34%).

Greater than 50% of patients with ulcer-related hemorrhage bleed without any preceding warning signs or symptoms.

PERFORATION The second most common ulcer-related complication is perforation, being reported in as many as 6–7% of PUD patients with an estimated 30-day mortality of >20%. Acute abdominal pain, tachycardia, and abdominal rigidity compose the classic triad associated with this complication. It is essential to remember that elderly patients or individuals who are immunosuppressed may not have this classic presentation. As in the case of bleeding, the incidence of perforation in the elderly appears to be increasing secondary to increased use of NSAIDs. Perforation of DUs has become less common in light of the increased rates of *H. pylori* eradication, with NSAID-induced GUs leading to perforation occurring more commonly. *Penetration* is a form of perforation in which the ulcer bed tunnels into an adjacent organ. DUs tend to penetrate posteriorly into the pancreas, leading to pancreatitis, whereas GUs tend to penetrate into the left hepatic lobe. Gastrocolic fistulas associated with GUs have also been described. Mortality for this complication can be >20% within 30 days.

GASTRIC OUTLET OBSTRUCTION Gastric outlet obstruction is the least common ulcer-related complication, occurring in 1–2% of patients. A patient may have relative obstruction secondary to ulcer-related inflammation and edema in the peripyloric and duodenal region. This process often resolves with ulcer healing. A fixed, mechanical obstruction secondary to scar formation in the peripyloric areas is also possible. The latter requires endoscopic (balloon dilation with or without placement of a biodegradable stent) or surgical intervention with a strictureplasty or gastrojejunostomy. Signs and symptoms relative to mechanical obstruction may develop insidiously. New onset of early satiety, nausea, vomiting, increase of postprandial abdominal pain, and weight loss should make gastric outlet obstruction a possible diagnosis.

Differential Diagnosis The list of GI and non-GI disorders that can mimic ulceration of the stomach or duodenum is quite extensive. The most commonly encountered diagnosis among patients seen for upper abdominal discomfort is functional dyspepsia (FD) or *essential dyspepsia*, which refers to a group of heterogeneous disorders typified by upper abdominal pain without the presence of an ulcer. The symptoms can range from postprandial fullness and early satiety to epigastric burning pain. The dichotomy of this symptom complex has led to the identification of two subcategories of FD including postprandial distress syndrome (PDS) and epigastric pain syndrome (EPS). Dyspepsia has been reported to occur in up to 30% of the U.S. population. Up to 80% of patients seeking medical care for dyspepsia have a negative diagnostic evaluation. The etiology of FD is not established, but recent studies suggest that postinfectious states, certain foods, and *H. pylori* infection may contribute to the pathogenesis of this common disorder.

Several additional disease processes that may present with “ulcer-like” symptoms include proximal GI tumors, gastroesophageal reflux, vascular disease, pancreaticobiliary disease (biliary colic, chronic pancreatitis), and gastroduodenal Crohn’s disease.

Diagnostic Evaluation In view of the poor predictive value of abdominal pain for the presence of a gastroduodenal ulcer and the multiple disease processes that can mimic this disease, the clinician is often confronted with having to establish the presence of an ulcer. Documentation of an ulcer requires either a radiographic (barium study, rarely done in today’s environment) or an endoscopic procedure. However, a large percentage of patients with symptoms suggestive of an ulcer have NUD; testing for *H. pylori* and antibiotic therapy (see below) are appropriate for individuals who are otherwise healthy and <45 years of age, before embarking on a diagnostic evaluation (Chap. 45).

Barium studies of the proximal GI tract are rarely used as a first test for documenting an ulcer. The sensitivity of older single-contrast barium meals for detecting a DU is as high as 80%, with a double-contrast study providing detection rates as high as 90%. Sensitivity for detection is decreased in small ulcers (<0.5 cm), with presence of previous scarring, or in postoperative patients. A DU appears as a well-demarcated crater, most often seen in the bulb (Fig. 324-10A). A GU may represent benign or malignant disease. Typically, a benign GU also appears as a discrete crater with radiating mucosal folds originating from the ulcer margin (Fig. 324-10B). Ulcers >3 cm in size or those associated with a mass are more often malignant. Unfortunately, up to 8% of GUs that appear to be benign by radiographic appearance are malignant by endoscopy or surgery. Radiographic studies that show a GU must be followed by endoscopy and biopsy.

Endoscopy provides the most sensitive and specific approach for examining the upper GI tract (Fig. 324-11). In addition to permitting direct visualization of the mucosa, endoscopy facilitates photographic documentation of a mucosal defect and tissue biopsy to rule out malignancy (GU) or *H. pylori*. Endoscopic examination is particularly helpful in identifying lesions too small to detect by radiographic examination, for evaluation of atypical radiographic abnormalities, or to determine if an ulcer is a source of blood loss.

Although the methods for diagnosing *H. pylori* are outlined in Chap. 163, a brief summary will be included here (Table 324-2). Several biopsy urease tests have been developed (PyloriTek, CLOtest, Hpfast, Pronto Dry) that have a sensitivity and specificity of >90–95%. Several noninvasive methods for detecting this organism have been developed. Three types of studies routinely used include serologic testing, the ¹³C- or ¹⁴C-urea breath test, and the fecal *H. pylori* (Hp) antigen test (monoclonal antibody test). A urinary Hp antigen test and a home breath test appear promising.

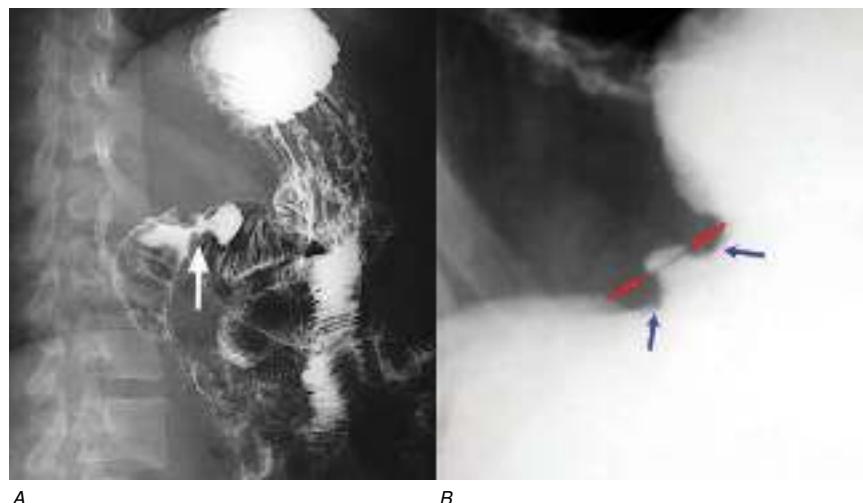


FIGURE 324-10 Barium study demonstrating (A) a benign duodenal ulcer and (B) a benign gastric ulcer.

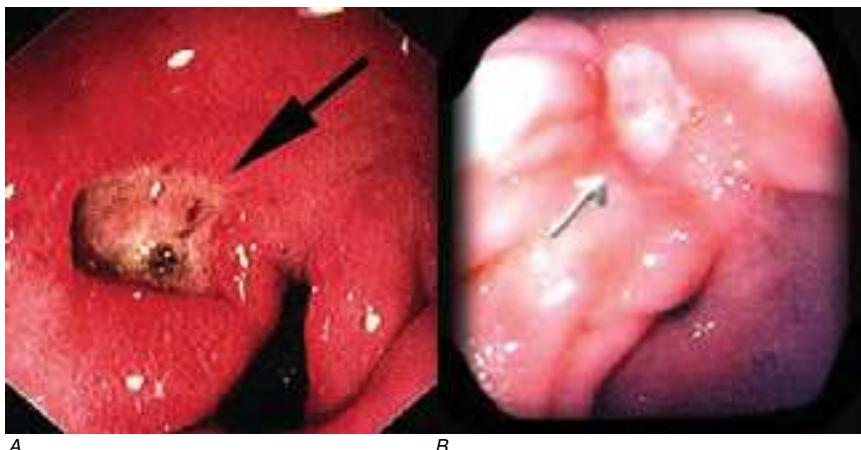


FIGURE 324-11 Endoscopy demonstrating (A) a benign duodenal ulcer and (B) a benign gastric ulcer.

Occasionally, specialized testing such as serum gastrin and gastric acid analysis may be needed in individuals with complicated or refractory PUD (see “Zollinger-Ellison Syndrome,” below). Screening for aspirin or NSAIDs (blood or urine) may also be necessary in refractory *H. pylori*-negative PUD patients.

TREATMENT

Peptic Ulcer Disease

Before the discovery of *H. pylori*, the therapy of PUD was centered on the old dictum of “no acid, no ulcer.” Although acid secretion is still important in the pathogenesis of PUD, eradication of *H. pylori* and therapy/prevention of NSAID-induced disease is the mainstay of treatment. A summary of commonly used drugs for treatment of acid peptic disorders is shown in Table 324-3.

ACID-NEUTRALIZING/INHIBITORY DRUGS

Antacids Before we understood the important role of histamine in stimulating parietal cell activity, neutralization of secreted acid with antacids constituted the main form of therapy for peptic ulcers. They are now rarely, if ever, used as the primary therapeutic agent but instead are often used by patients for symptomatic relief of dyspepsia. The most commonly used agents are mixtures of aluminum hydroxide and magnesium hydroxide. Aluminum hydroxide can produce constipation and phosphate depletion; magnesium hydroxide may cause loose stools. Many of the commonly used

antacids (e.g., Maalox, Mylanta) have a combination of both aluminum and magnesium hydroxide in order to avoid these side effects. The magnesium-containing preparation should not be used in chronic renal failure patients because of possible hypermagnesemia, and aluminum may cause chronic neurotoxicity in these patients.

Calcium carbonate and sodium bicarbonate are potent antacids with varying levels of potential problems. The long-term use of calcium carbonate (converts to calcium chloride in the stomach) can lead to milk-alkali syndrome (hypercalcemia and hyperphosphatemia with possible renal calcinosis and progression to renal insufficiency). Sodium bicarbonate may induce systemic alkalosis.

***H₂* Receptor Antagonists** Four of these agents are presently available (cimetidine, ranitidine, famotidine, and nizatidine), and their structures share homology with histamine. Although each has different potency, all will significantly inhibit basal and stimulated acid secretion to comparable levels when used at therapeutic doses. Moreover, similar ulcer-healing rates are achieved with each drug when used at the correct dosage. Presently, this class of drug is often used for treatment of active ulcers (4–6 weeks) in combination with antibiotics directed at eradicating *H. pylori* (see below).

Cimetidine was the first *H₂* receptor antagonist used for the treatment of acid peptic disorders. Cimetidine may have weak

TABLE 324-3 Drugs Used in the Treatment of Peptic Ulcer Disease

DRUG TYPE/MECHANISM	EXAMPLES	DOSE
Acid-Suppressing Drugs		
Antacids	Mylanta, Maalox, Tums, Gaviscon	100–140 meq/L 1 and 3 h after meals and hs
<i>H₂</i> receptor antagonists	Cimetidine	400 mg bid
	Ranitidine	300 mg hs
	Famotidine	40 mg hs
	Nizatidine	300 mg hs
Proton pump inhibitors	Omeprazole	20 mg/d
	Lansoprazole	30 mg/d
	Rabeprazole	20 mg/d
	Pantoprazole	40 mg/d
	Esomeprazole	20 mg/d
	Dexlansoprazole	30 mg/d
Mucosal Protective Agents		
Sucralfate	Sucralfate	1 g qid
Prostaglandin analogue	Misoprostol	200 µg qid
Bismuth-containing compounds	Bismuth subsalicylate (BSS)	See anti- <i>H. pylori</i> regimens (Table 324-4)

TABLE 324-2 Tests for Detection of *Helicobacter pylori*

TEST	SENSITIVITY/ SPECIFICITY, %	COMMENTS
Invasive (Endoscopy/Biopsy Required)		
Rapid urease	80–95/95–100	Simple, false negative with recent use of PPIs, antibiotics, or bismuth compounds
Histology	80–90/>95	Requires pathology processing and staining; provides histologic information
Culture	—/—	Time-consuming, expensive, dependent on experience; allows determination of antibiotic susceptibility
Noninvasive		
Serology	>80/>90	Inexpensive, convenient; not useful for early follow-up
Urea breath test	>90/>90	Simple, rapid; useful for early follow-up; false negatives with recent therapy (see rapid urease test); exposure to low-dose radiation with ¹⁴ C test
Stool antigen	>90/>90	Inexpensive, convenient

Abbreviation: PPIs, proton pump inhibitors.

Abbreviation: hs, at bedtime (*hora somni*).

antiandrogenic side effects resulting in reversible gynecomastia and impotence, primarily in patients receiving high doses for prolonged periods of time (months to years). In view of cimetidine's ability to inhibit cytochrome P450, careful monitoring of drugs such as warfarin, phenytoin, and theophylline is indicated with long-term usage. Other rare reversible adverse effects reported with cimetidine include confusion and elevated levels of serum aminotransferases, creatinine, and serum prolactin. Ranitidine, famotidine, and nizatidine are more potent H₂ receptor antagonists than cimetidine. Each can be used once a day at bedtime for ulcer prevention, which was commonly done before the discovery of *H. pylori* and the development of proton pump inhibitors (PPIs). Patients may develop tolerance to H₂ blockers, a rare event with PPIs (see below). Comparable nighttime dosing regimens are cimetidine 800 mg, ranitidine 300 mg, famotidine 40 mg, and nizatidine 300 mg.

Additional rare, reversible systemic toxicities reported with H₂ receptor antagonists include pancytopenia, neutropenia, anemia, and thrombocytopenia, with a prevalence rate varying from 0.01 to 0.2%. Cimetidine and ranitidine (to a lesser extent) can bind to hepatic cytochrome P450; famotidine and nizatidine do not. Ranitidine and nizatidine were taken off of the market due to contamination of the drug with *N*-nitrosodimethylamine (NDMA), a known carcinogen.

Proton Pump (H⁺,K⁺-ATPase) Inhibitors Omeprazole, esomeprazole, lansoprazole, rabeprazole, and pantoprazole are substituted benzimidazole derivatives that covalently bind and irreversibly inhibit H⁺,K⁺-ATPase. Esomeprazole is the S-enantiomer of omeprazole, which is a racemic mixture of both S- and R-optical isomers. The R-isomer of lansoprazole, dexlansoprazole, is the most recent PPI approved for clinical use. Its reported advantage is a dual delayed-release system aimed at improving treatment of gastroesophageal reflux disease (GERD). These are the most potent acid inhibitory agents available. Omeprazole and lansoprazole are the PPIs that have been used for the longest time. Both are acid-labile and are administered as enteric-coated granules in a sustained-release capsule that dissolves within the small intestine at a pH of 6. Lansoprazole is available in an orally disintegrating tablet that can be taken with or without water, an advantage for individuals who have significant dysphagia. Absorption kinetics are similar to the capsule. In addition, a lansoprazole-naproxen combination preparation that has been made available is targeted at decreasing NSAID-related GI injury (see below). Omeprazole is available as non-enteric-coated granules mixed with sodium bicarbonate in a powder form that can be administered orally or via gastric tube. The sodium bicarbonate has two purposes: to protect the omeprazole from acid degradation and to promote rapid gastric alkalinization and subsequent proton pump activation, which facilitates rapid action of the PPI. Pantoprazole and rabeprazole are available as enteric-coated tablets. Pantoprazole is also available as a parenteral formulation for intravenous use. These agents are lipophilic compounds; upon entering the parietal cell, they are protonated and trapped within the acid environment of the tubulovesicular and canalicular system. These agents potently inhibit all phases of gastric acid secretion. Onset of action is rapid, with a maximum acid inhibitory effect between 2 and 6 h after administration and duration of inhibition lasting up to 72–96 h. With repeated daily dosing, progressive acid inhibitory effects are observed, with basal and secretagogue-stimulated acid production being inhibited by >95% after 1 week of therapy. The half-life of PPIs is ~18 h; thus, it can take between 2 and 5 days for gastric acid secretion to return to normal levels once these drugs have been discontinued. Because the pumps need to be activated for these agents to be effective, their efficacy is maximized if they are administered before a meal (except for the immediate-release formulation of omeprazole) (e.g., in the morning before breakfast). Mild to moderate hypergastrinemia has been observed in patients taking these drugs. Carcinoid tumors developed in some animals given the drugs preclinically; however, extensive experience has failed to demonstrate gastric carcinoid tumor development in humans. Serum gastrin levels return to

normal levels within 1–2 weeks after drug cessation. Rebound gastric acid hypersecretion has been described in *H. pylori*-negative individuals after discontinuation of PPIs. It occurs even after relatively short-term usage (2 months) and may last for up to 2 months after the PPI has been discontinued. The mechanism involves gastrin-induced hyperplasia and hypertrophy of histamine-secreting ECL cells. The clinical relevance of this observation is that individuals may have worsening symptoms of GERD or dyspepsia upon stopping the PPI. Gradual tapering of the PPI and switching to an H₂ receptor antagonist may prevent this from occurring. *H. pylori*-induced inflammation and concomitant decrease in acid production may explain why this does not occur in *H. pylori*-positive patients. IF production is also inhibited, but vitamin B₁₂ deficiency anemia is uncommon, probably because of the large stores of the vitamin. As with any agent that leads to significant hypochlorhydria, PPIs may interfere with absorption of drugs such as ketoconazole, ampicillin, iron, and digoxin. Hepatic cytochrome P450 can be inhibited by the earlier PPIs (omeprazole, lansoprazole). Rabeprazole, pantoprazole, and esomeprazole do not appear to interact significantly with drugs metabolized by the cytochrome P450 system. The overall clinical significance of this observation is not definitely established. Caution should be taken when using theophylline, warfarin, diazepam, atazanavir, and phenytoin concomitantly with PPIs.

The list of potential side effects with long-term PPI use has steadily grown over the years. These agents are commonly used since several formulations have become available as over-the-counter medications. Moreover, up to 70% of current prescriptions for long-term PPIs may be unwarranted and between 35 and 60% of in-hospital use of PPIs may be inappropriate. Interpretation of the multiple studies should take into consideration that the vast majority were retrospective observational studies in which confounding factors could not be accounted for entirely.

Long-term acid suppression, especially with PPIs, has been associated with a higher incidence of community-acquired pneumonia as well as community- and hospital-acquired *Clostridium difficile*-associated disease. A meta-analysis showed a 74% increased risk of *C. difficile* infection and a 2.5-fold higher risk of reinfection as compared to nonusers. In light of these concerns, the FDA published a safety alert regarding the association between *C. difficile* infection and PPI use. Although the risk of spontaneous bacterial peritonitis in cirrhotics was thought to be increased, the data here are less supportive. The impact of PPI-induced changes in the host microbiome is postulated to play a role in the increased risk of infection, but this theory needs to be confirmed. These observations require confirmation but should alert the practitioner to take caution when recommending these agents for long-term use, especially in elderly patients at risk for developing pneumonia or *C. difficile* infection.

Diarrhea is also associated with PPI use, which in some cases has been associated with the development of collagenous colitis (hazard ratio of 4.5), particularly with lansoprazole. The mechanism for PPI-induced collagenous colitis is unclear, but in vitro studies demonstrate that PPIs may induce collagen gene expression. The colitis usually resolves with cessation of the PPI.

A population-based study revealed that long-term use of PPIs was associated with the development of hip fractures in older women. The absolute risk of fracture remained low and may be zero despite an observed increase associated with the dose and duration of acid suppression. The mechanism for this observation is not clear, and this finding must be confirmed before making broad recommendations regarding the discontinuation of these agents in patients who benefit from them. Long-term use of PPIs has also been implicated in the development of iron, vitamin B₁₂, and magnesium deficiency. A meta-analysis of nine observational studies found a 40% increase in hypomagnesemia in PPI users as compared to nonusers. One approach to consider in patients needing to take PPIs long term is to check a complete blood count looking for evidence of anemia due to iron or vitamin B₁₂ deficiency, vitamin B₁₂ level, and a magnesium level after 1–2 years of PPI use, but these recommendations are not evidence based or recommended by expert opinion. PPIs may exert

a negative effect on the antiplatelet effect of clopidogrel. Although the evidence is mixed and inconclusive, a small increase in mortality and readmission rate for coronary events was seen in patients receiving a PPI while on clopidogrel in earlier studies. Subsequently, three meta-analyses reported an inverse correlation between clopidogrel and PPI use; therefore, the influence of this drug interaction on mortality is not clearly established. The mechanism involves the competition of the PPI and clopidogrel with the same cytochrome P450 (CYP2C19). Whether this is a class effect of PPIs is unclear; there appears to be at least a theoretical advantage of pantoprazole over the other PPIs, but this has not been confirmed. This drug interaction is particularly relevant in light of the common use of aspirin and clopidogrel for prevention of coronary events and the efficacy of PPIs in preventing GI bleeding in these patients. The FDA has made several recommendations while awaiting further evidence to clarify the impact of PPI therapy on clopidogrel use. Health care providers should continue to prescribe clopidogrel to patients who require it and should reevaluate the need for starting or continuing treatment with a PPI. From a practical standpoint, additional recommendations to consider include the following: Patients taking clopidogrel with aspirin, especially with other GI risk factors for bleeding, should receive GI protective therapy. Although high-dose H₂ blockers have been considered an option, these do not appear to be as effective as PPIs. If PPIs are to be given, some have recommended that there be a 12-h separation between administration of the PPI and clopidogrel to minimize competition of the two agents with the involved cytochrome P450. One option is to give the PPI 30 min before breakfast and the clopidogrel at bedtime. Insufficient data are available to firmly recommend one PPI over another. Additional concerning side effects with long-term PPI use include increased cardiac risks independent of clopidogrel use, dementia, and acute and chronic kidney injury. Again, the data are often retrospective and confounding variables were not consistently eliminated, thus making it difficult to establish a definitive association between PPIs and the toxicities outlined. A summary of the side effects with the corresponding relative risks is shown in Fig. 324-12. Ultimately, heightened awareness of inappropriate long-term use of PPIs is paramount. Patients aged ≥65 years of age have a higher risk for some of the long-term side effects of PPIs highlighted above, in part due to the higher prevalence of concomitant chronic diseases. It is therefore essential to carefully select individuals, especially among the elderly, who need long-term PPI therapy and discontinue it in those individuals who do not need it. Abrupt withdrawal of a PPI in a long-term user may result in a component of rebound hyperacidity; thus, this agent should be tapered gradually over the course of 1–2 weeks with possible transition to an H₂ blocker for a short period of time.

Development of novel acid inhibitory agents continues in an attempt to primarily address the need for better agents to treat

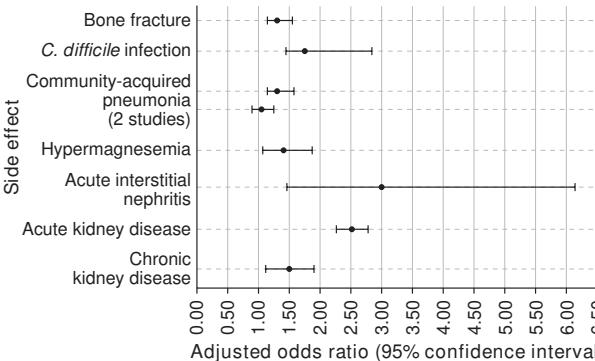


FIGURE 324-12. Evidence supporting the potential adverse effects of proton pump inhibitor drugs. (Adapted from AJ Schoenfeld, D Grady: Adverse effects associated with proton pump inhibitors. *JAMA Intern Med* 176:172, 2016.)

GERD. For example, modified H₂ blockers with greater potency and duration as well as novel PPIs with longer half-life and potency are under study. For example, tenatoprazole is a PPI containing an imidazopyridine ring instead of a benzimidazole ring, which promotes irreversible proton pump inhibition. This agent has a longer half-life than the other PPIs and may be beneficial for inhibiting nocturnal acid secretion, which has significant relevance in GERD. Additional PPIs with longer half-life and combined with other agents are being studied, but the details are beyond the scope of this chapter. A second new class of agents is the potassium-competitive acid pump antagonists (P-CAPs). These compounds inhibit gastric acid secretion via potassium competitive binding of the H⁺,K⁺-AT-Pase. Revaprazan, vonoprazan and tegoprazan are agents approved for use in Korea and Japan. Vonoprazan may be superior to PPIs when combined with antibiotics for the treatment of *H. pylori*, and this novel agent has been awarded Fast Track status by the FDA for the treatment of *H. pylori* in combination with both amoxicillin and clarithromycin and with amoxicillin alone.

CYTOPROTECTIVE AGENTS

Sucralfate Sucralfate is a complex sucrose salt in which the hydroxyl groups have been substituted by aluminum hydroxide and sulfate. This compound is insoluble in water and becomes a viscous paste within the stomach and duodenum, binding primarily to sites of active ulceration. Sucralfate may act by several mechanisms: serving as a physicochemical barrier, promoting a trophic action by binding growth factors such as EGF, enhancing prostaglandin synthesis, stimulating mucus and bicarbonate secretion, and enhancing mucosal defense and repair. Toxicity from this drug is rare, with constipation being most common (2–3%). It should be avoided in patients with chronic renal insufficiency to prevent aluminum-induced neurotoxicity. Hypophosphatemia and gastric bezoar formation have also been reported rarely. Standard dosing of sucralfate is 1 g qid.

Bismuth-Containing Preparations Sir William Osler considered bismuth-containing compounds the drug of choice for treating PUD. The resurgence in the use of these agents is due to their effect against *H. pylori*. Colloidal bismuth subcitrate (CBS) and bismuth subsalicylate (BSS; Pepto-Bismol) are the most widely used preparations. The mechanism by which these agents induce ulcer healing is unclear. Adverse effects with short-term use include black stools, constipation, and darkening of the tongue. Long-term use with high doses, especially with the avidly absorbed CBS, may lead to neurotoxicity. These compounds are commonly used as one of the agents in an anti-*H. pylori* regimen (see below).

Prostaglandin Analogues In view of their central role in maintaining mucosal integrity and repair, stable prostaglandin analogues were developed for the treatment of PUD. The mechanism by which this rapidly absorbed drug provides its therapeutic effect is through enhancement of mucosal defense and repair. The most common toxicity noted with this drug is diarrhea (10–30% incidence). Other major toxicities include uterine bleeding and contractions; misoprostol is contraindicated in women who may be pregnant, and women of childbearing age must be made clearly aware of this potential drug toxicity. The standard therapeutic dose is 200 µg qid.

Miscellaneous Drugs A number of drugs including anticholinergic agents and tricyclic antidepressants were used for treating acid peptic disorders, but in light of their toxicity and the development of potent antisecretory agents, these are rarely, if ever, used today. Newer agents such as teprenone, an acyclic polyisoprenoid compound used as a gastric mucosal protector that is employed to treat gastritis and GUs outside of the United States; plant-based therapies; and CCK2 receptor antagonists are intriguing therapies but require further evaluation.

THERAPY OF *H. PYLORI*

The physician's goal in treating PUD is to provide relief of symptoms (pain or dyspepsia), promote ulcer healing, and ultimately

prevent ulcer recurrence and complications. The greatest influence of understanding the role of *H. pylori* in peptic disease has been the ability to prevent recurrence. Documented eradication of *H. pylori* in patients with PUD is associated with a dramatic decrease in ulcer recurrence to <10–20% as compared to 59% in GU patients and 67% in DU patients when the organism is not eliminated. Eradication of the organism may lead to diminished recurrent ulcer bleeding. The effect of its eradication on ulcer perforation is unclear.

Extensive effort has been made in determining who of the many individuals with *H. pylori* infection should be treated. The common conclusion arrived at by multiple consensus conferences around the world is that *H. pylori* should be eradicated in patients with documented PUD. This holds true independent of time of presentation (first episode or not), severity of symptoms, presence of confounding factors such as ingestion of NSAIDs, or whether the ulcer is in remission. Some have advocated treating patients with a history of documented PUD who are found to be *H. pylori* positive by stool antigen or breath testing. Between 60 and 90% of patients with gastric MALT lymphoma experience complete remission of the tumor in response to *H. pylori* eradication. The Maastricht IV/Florence Consensus Report recommends a test-and-treat approach for patients with uninvestigated dyspepsia if the local incidence of *H. pylori* is >20%. The American College of Gastroenterology (ACG) clinical guidelines (developed for North America) recommend that individuals aged <60 years with uninvestigated dyspepsia should be tested and treated for *H. pylori*. In addition, recommendations from this consensus report and the ACG clinical guidelines include testing and offering eradication of *H. pylori* in patients who will be using NSAIDs (including low-dose aspirin) on a long-term basis, especially if there is a prior history of PUD. These individuals will require continued PPI treatment as well as eradication treatment, because eradication of the organism alone does not eliminate the risk of gastroduodenal ulcers in patients already receiving long-term NSAIDs. Treating patients with NUD to prevent gastric cancer or patients with GERD requiring long-term acid suppression remains controversial. Guidelines from the ACG suggest eradication of *H. pylori* in patients who have undergone resection of early gastric cancer. The Maastricht IV/Florence Consensus Report also evaluated *H. pylori* treatment in gastric cancer prevention and recommends that eradication should be considered in the following situations: first-degree relatives of family members with gastric cancer; patients with previous gastric neoplasm treated by endoscopic or subtotal resection; individuals with a risk of gastritis (severe pan-gastritis or body-predominant gastritis) or severe atrophy; patients with gastric acid inhibition for >1 year; individuals with strong environmental risk factors for gastric cancer (heavy smoking; high exposure to dust, coal, quartz, or cement; and/or work in quarries); and *H. pylori*-positive patients with a fear of gastric cancer. Finally, the ACG clinical guidelines recommend testing and offering *H. pylori* eradication to patients with unexplained iron deficiency anemia and idiopathic thrombocytopenic purpura. Despite this, concerns have been raised about the widespread use of antibiotics for the therapy of all cases of *H. pylori* positivity, including the potential for increased bacterial resistance rates, reported weight gain, and alteration of the microbiome.

Multiple drugs have been evaluated in the therapy of *H. pylori*. No single agent is effective in eradicating the organism. Combination therapy for 14 days provides the greatest efficacy, although regimens based on sequential administration of antibiotics also appear promising (see below). A shorter administration course (7–10 days), although attractive, has not proved as successful as the 14-day regimens. The agents used with the greatest frequency include amoxicillin, metronidazole, tetracycline, clarithromycin, and bismuth compounds.

Suggested treatment regimens for *H. pylori* are outlined in Table 324-4. Choice of a particular regimen will be influenced by several factors, including efficacy, patient tolerance, existing antibiotic resistance, prior antibiotic use, and cost of the drugs. The aim for initial eradication rates should be 85–90%. Dual therapy

(PPI plus amoxicillin, PPI plus clarithromycin, ranitidine bismuth citrate [Tritec] plus clarithromycin) is not recommended in view of studies demonstrating eradication rates of <80–85%. The combination of bismuth, metronidazole, and tetracycline was the first triple regimen found effective against *H. pylori*. The combination of two antibiotics plus either a PPI, H₂ blocker, or bismuth compound has comparable success rates. Addition of acid suppression assists in providing early symptom relief and enhances bacterial eradication.

Triple therapy, although effective, has several drawbacks, including the potential for poor patient compliance and drug-induced side effects. Compliance is being addressed by simplifying the regimens so that patients can take the medications twice a day. Simpler (dual therapy) and shorter regimens (7 and 10 days) are not as effective as triple therapy for 14 days. Two anti-*H. pylori* regimens are available in prepackaged formulation: Prevpac (lansoprazole, clarithromycin, and amoxicillin) and Helidac (BSS, tetracycline, and metronidazole). The contents of the Prevpac are to be taken twice per day for 14 days, whereas Helidac constituents are taken four times per day with an antisecretory agent (PPI or H₂ blocker), also for at least 14 days. Clarithromycin-based triple therapy should be avoided in settings where *H. pylori* resistance to this agent exceeds 15%.

Side effects have been reported in up to 20–30% of patients on triple therapy. Bismuth may cause black stools, constipation, or darkening of the tongue. The most feared complication with amoxicillin is pseudomembranous colitis, but this occurs in <1–2% of patients. Amoxicillin can also lead to antibiotic-associated diarrhea, nausea, vomiting, skin rash, and allergic reaction. Concomitant use of probiotics may ameliorate some of the antibiotic side effects (see below). Tetracycline has been reported to cause rashes and, very rarely, hepatotoxicity and anaphylaxis.

One important concern with treating patients who may not need therapy is the potential for development of antibiotic-resistant strains. The incidence and type of antibiotic-resistant *H. pylori* strains vary worldwide. Strains resistant to metronidazole, clarithromycin, amoxicillin, and tetracycline have been described, with the latter two being uncommon. Antibiotic-resistant strains are the most common cause for treatment failure in compliant patients. Unfortunately, *in vitro* resistance does not predict outcome in patients. Culture and sensitivity testing of *H. pylori* is not performed routinely. Although resistance to metronidazole has been found in as many as 30% of isolates in North America and 80% in developing countries, triple therapy is effective in eradicating the organism in >50% of patients infected with a resistant strain. Clarithromycin resistance is seen in 13–16% of individuals in the United States, with resistance to amoxicillin being <1% and resistance to both metronidazole and clarithromycin in the 5% range. Resistance to tetracycline and rifabutin (see below) is reported to be <2% in the United States. In light of the paucity of *H. pylori* antibiotic real-time resistance data, asking the patient about prior antibiotic exposure should be included in the decision-making and used as a surrogate for potential antibiotic resistance, especially when it comes to prior macrolide use. Clarithromycin use should be excluded in patients with prior macrolide usage. An approach to antibiotic selection for *H. pylori* therapy has been recommended in the ACG clinical guidelines (Fig. 324-13).

Failure of *H. pylori* eradication with triple therapy in a compliant patient is usually due to infection with a resistant organism. A series of salvage therapies for *H. pylori* are shown in Table 324-5. Quadruple therapy (Table 324-4), where clarithromycin is substituted for metronidazole (or vice versa), should be the next step. The combination of PPI, amoxicillin, and rifabutin for 10 days has also been used successfully (86% cure rate) in patients infected with resistant strains. Additional regimens considered for second-line therapy include levofloxacin-based triple therapy (levofloxacin, amoxicillin, PPI) for 10 days and furazolidone-based triple therapy (furazolidone, amoxicillin, PPI) for 14 days. Unfortunately, there is no universally accepted treatment regimen recommended for patients in whom two courses of antibiotics have failed. If eradication is still not achieved in a compliant patient, then culture and sensitivity of the

TABLE 324-4 Recommended First-Line Therapies for *H. pylori* Infection

REGIMEN	DRUGS (DOSES)	DOSING FREQUENCY	DURATION (DAYS)	FDA APPROVAL
Clarithromycin triple	PPI (standard or double dose)	bid	14	Yes ^a
	Clarithromycin (500 mg)			
	Amoxicillin (1 g) or metronidazole (500 mg tid)			
Bismuth quadruple	PPI (standard dose)	bid	10–14	No ^b
	Bismuth subcitrate (120–300 mg) or subsalicylate (300 mg)	qid		
	Tetracycline (500 mg)	qid		
	Metronidazole (250–500 mg)	qid (250 mg) tid to qid (500 mg)		
Concomitant	PPI (standard dose)	bid	10–14	No
	Clarithromycin (500 mg)			
	Amoxicillin (1 g)			
	Nitroimidazole (500 mg) ^c			
Sequential	PPI (standard dose)	bid	5–7	No
	PPI, clarithromycin (500 mg) + nitroimidazole (500 mg) ^c	bid	5–7	
Hybrid	PPI (standard dose) + amoxicillin (1 g)	bid	7	No
	PPI, amoxicillin, clarithromycin (500 mg), nitroimidazole (500 mg) ^c	bid	7	
Levofloxacin triple	PPI (standard or double dose) + amoxicillin (1 g)	bid	5–7	No
	Levofloxacin (500 mg)	qd		
	Amoxicillin (1 g)	bid		
Levofloxacin sequential	PPI (standard or double dose) + amoxicillin (1 g)	bid	5–7	No
	PPI, amoxicillin, levofloxacin (500 mg qd), nitroimidazole (500 mg) ^c	bid	5–7	
LOAD	Levofloxacin (250 mg)	qd	7–10	No
	PPI (double dose)	qd		
	Nitazoxanide (500 mg)	bid		
	Doxycycline (100 mg)	qd		

^aSeveral PPI, clarithromycin, and amoxicillin combinations have achieved FDA approval. The regimen of a PPI, clarithromycin, and metronidazole is not an FDA-approved treatment regimen. ^bThe regimen of a PPI, bismuth, tetracycline, and metronidazole combined with a PPI for 10 days is an FDA-approved treatment regimen. ^cMetronidazole or tinidazole.

Abbreviations: bid, twice daily; FDA, Food and Drug Administration; PPI, proton pump inhibitor; tid, three times daily; qd, once daily; qid, four times daily.

Source: Reproduced with permission from WD Chey et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. Am J Gastroenterol 112:212, 2017.

organism should be considered. One challenge with this approach is that culture and sensitivity testing is cumbersome and not widely available; thus, *H. pylori* resistance data within specific communities are often not available. Non-culture-based approaches using molecular markers to determine potential resistance through stool testing are being developed but are not widely available. Additional factors that may lower eradication rates include the patient's country of origin (higher in Northeast Asia than other parts of Asia or Europe) and cigarette smoking. In addition, meta-analysis suggests that even the most effective regimens (quadruple therapy including PPI, bismuth, tetracycline, and metronidazole and triple therapy including PPI, clarithromycin, and amoxicillin) may have suboptimal eradication rates (<80%), thus demonstrating the need for the development of more efficacious treatments.

In view of the observation that 15–25% of patients treated with first-line therapy may still remain infected with the organism, new approaches to treatment have been explored. One promising approach is sequential therapy. Regimens examined consist of 5 days of amoxicillin and a PPI, followed by an additional 5 days of PPI plus tinidazole and clarithromycin or levofloxacin. One promising regimen that has the benefit of being shorter in duration, easier to take, and less expensive is 5 days of concomitant therapy (PPI twice daily, amoxicillin 1 g twice daily, levofloxacin 500 mg twice daily, and tinidazole 500 mg twice daily). Initial studies have demonstrated eradication rates of >90% with good patient tolerance. Confirmation of these findings and applicability of this approach in the United States are needed, although some experts are recommending abandoning clarithromycin-based triple therapy in the United States for the concomitant therapy or the alternative sequential therapies highlighted above.

Innovative non-antibiotic-mediated approaches have been explored in an effort to improve eradication rates of *H. pylori*. Pretreatment of patients with *N*-acetylcysteine as a mucolytic agent to destroy the *H. pylori* biofilm and therefore impair antibiotic resistance has been examined, but more studies are needed to confirm the applicability of this approach. In vitro studies suggest that certain probiotics like *Lactobacillus* or its metabolites can inhibit *H. pylori*. Administration of probiotics has been attempted in several clinical studies in an effort to maximize antibiotic-mediated eradication with varying results. Overall, it appears that the use of certain probiotics, such as *Lactobacillus* spp., *Saccharomyces* spp., *Bifidobacterium* spp., and *Bacillus clausii*, did not alter eradication rates but importantly decreased antibiotic-associated side effects including nausea, dysgeusia, diarrhea, and abdominal discomfort/pain, resulting in enhanced tolerability of *H. pylori* therapies. Additional studies are needed to confirm the potential benefits of probiotics in this setting. Statins, specifically atorvastatin, have been used with some success as an adjunct to quadruple therapy in patients with NUD.

Reinfection after successful eradication of *H. pylori* is rare in the United States (<1% per year). If recurrent infection occurs within the first 6 months after completing therapy, the most likely explanation is recrudescence as opposed to reinfection.

THERAPY OF NSAID-RELATED GASTRIC OR DUODENAL INJURY

Medical intervention for NSAID-related mucosal injury includes treatment of an active ulcer and primary prevention of future injury. Recommendations for the treatment and primary prevention of NSAID-related mucosal injury are listed in Table 324-6.

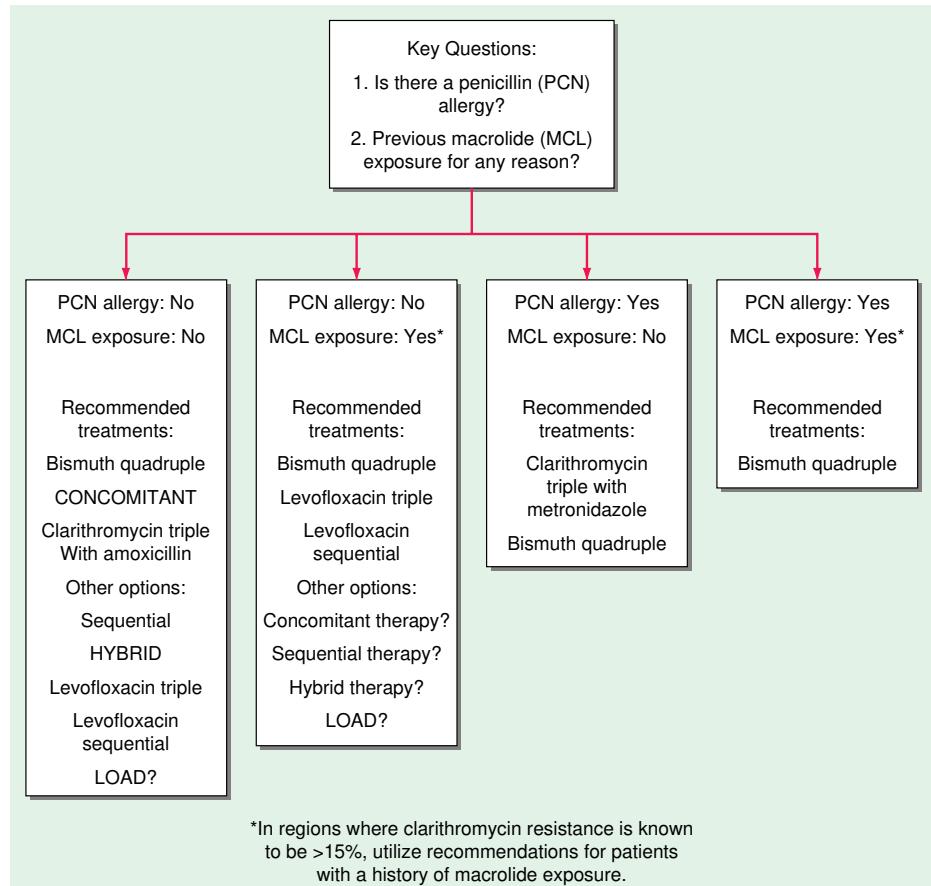


FIGURE 324-13 Approach to selecting antibiotics for patients with *H. pylori* infection. LOAD, levofloxacin, omeprazole, nitazoxanide, and doxycycline. (Reproduced with permission from WD Chey et al: ACG clinical guideline: Treatment of Helicobacter pylori infection. *Am J Gastroenterol* 112:212, 2017.)

TABLE 324-5 Salvage Therapies for *H. pylori* Infection

REGIMEN	DRUGS (DOSES)	DOSING FREQUENCY	DURATION (DAYS)	FDA APPROVAL
Bismuth quadruple	PPI (standard dose) Bismuth subcitrate (120–300 mg) or subsalicylate (300 mg) Tetracycline (500 mg) Metronidazole (500 mg)	bid qid qid tid or qid	14	No ^a
Levofloxacin triple	PPI (standard dose) Levofloxacin (500 mg) Amoxicillin (1 g)	bid qd bid	14	No
Concomitant	PPI (standard dose) Clarithromycin (500 mg) Amoxicillin (1 g) Nitroimidazole (500 mg)	bid bid bid bid or tid	10–14	No
Rifabutin triple	PPI (standard dose) Rifabutin (300 mg) Amoxicillin (1 g)	bid qd bid	10	No
High-dose dual	PPI (standard to double dose) Amoxicillin (1 g tid or 750 mg qid)	tid or qid tid or qid	14	No

^aPPI, bismuth, tetracycline, and metronidazole prescribed separately is not an FDA-approved treatment regimen. However, Pylera, a combination product containing bismuth subcitrate, tetracycline, and metronidazole, combined with a PPI for 10 days is an FDA-approved treatment regimen.

Abbreviations: bid, twice daily; FDA, Food and Drug Administration; PPI, proton pump inhibitor; tid, three times daily; qd, once daily; qid, four times daily.

Source: Reproduced with permission from WD Chey et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. *Am J Gastroenterol* 112:212, 2017.

TABLE 324-6 Recommendations for Treatment of NSAID-Related Mucosal Injury

CLINICAL SETTING	RECOMMENDATION
Active ulcer	
NSAID discontinued	H ₂ receptor antagonist or PPI
NSAID continued	PPI
Prophylactic therapy	Misoprostol PPI Selective COX-2 inhibitor
<i>H. pylori</i> infection	Eradication if active ulcer present or there is a past history of peptic ulcer disease

Abbreviations: COX-2, isoenzyme of cyclooxygenase; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

Ideally, the injurious agent should be stopped as the first step in the therapy of an active NSAID-induced ulcer. If that is possible, then treatment with one of the acid inhibitory agents (H₂ blockers, PPIs) is indicated. Cessation of NSAIDs is not always possible because of the patient's severe underlying disease. Only PPIs can heal GUs or DUs, independent of whether NSAIDs are discontinued.

The widespread use of NSAIDs has created some concern due to the increasing likelihood of GI and CV side effects associated with these agents. The approach to primary prevention has included avoiding the agent, using the lowest possible dose of the agent for the shortest period of time possible, using NSAIDs that are theoretically less injurious, using newer topical NSAID preparations, and/or using concomitant medical therapy to prevent NSAID-induced injury. Several nonselective NSAIDs that are associated with a lower likelihood of GI and CV toxicity include naproxen and ibuprofen, although the beneficial effect may be eliminated if higher dosages of the agents are used. Primary prevention of NSAID-induced ulceration can be accomplished by a PPI and, if not tolerated, misoprostol (200 µg qid). High-dose H₂ blockers (famotidine 40 mg bid) have also shown some promise in preventing endoscopically documented ulcers, although PPIs are superior. The highly selective COX-2 inhibitors, celecoxib and rofecoxib, are 100 times more selective inhibitors of COX-2 than standard NSAIDs, leading to gastric or duodenal mucosal injury that is comparable to placebo; their utilization led to an increase in CV events and withdrawal from the market. Additional caution was engendered when the CLASS study demonstrated that the advantage of celecoxib in preventing GI complications was offset when low-dose aspirin was used simultaneously. Therefore, gastric protection therapy is required in individuals taking COX-2 inhibitors and aspirin prophylaxis. Finally, much of the work demonstrating the benefit of COX-2 inhibitors and PPIs on GI injury has been performed in individuals of average risk; it is unclear if the same level of benefit will be achieved in high-risk patients. For example, concomitant use of warfarin and a COX-2 inhibitor was associated with rates of GI bleeding similar to those observed in patients taking nonselective NSAIDs. A combination of factors, including withdrawal of the majority of COX-2 inhibitors from the market, the observation that low-dose aspirin appears to diminish the beneficial effect of COX-2-selective inhibitors, and the growing use of aspirin for prophylaxis of CV events, has significantly altered the approach to gastric protective therapy during the use of NSAIDs. A set of guidelines for the approach to the use of NSAIDs was published by the ACG and is shown in Table 324-7. Individuals who are not at risk for CV events, do not use aspirin, and are without risk for GI complications can receive nonselective NSAIDs without gastric protection. In those without CV risk factors but with a high potential risk (prior GI bleeding or multiple GI risk factors) for NSAID-induced GI toxicity, cautious use of a selective COX-2 inhibitor and co-therapy with high-dose PPI or misoprostol are recommended. Individuals at moderate GI risk without cardiac risk factors can be treated with a COX-2 inhibitor alone or with a nonselective NSAID with PPI or misoprostol. Individuals with CV risk factors, who require low-dose

TABLE 324-7 Guide to NSAID Therapy

	NO/LOW NSAID GI RISK	NSAID GI RISK
No CV risk (no aspirin)	Traditional NSAID	Coxib or Traditional NSAID + PPI or misoprostol Consider non-NSAID therapy
CV risk (consider aspirin)	Traditional NSAID + PPI or misoprostol if GI risk warrants gastroprotection Consider non-NSAID therapy	A gastroprotective agent must be added if a traditional NSAID is prescribed Consider non-NSAID therapy

Abbreviations: CV, cardiovascular; GI, gastrointestinal; NSAID, nonsteroidal anti-inflammatory drug; PPI, proton pump inhibitor.

Source: Republished with permission of MJH Life Sciences, LLC, from COX-2 inhibitor use after Vioxx: careful balance orend of the rope?, Fendrick AM, 10(11 Pt 1): 2004; permission conveyed through Copyright Clearance Center, Inc.

aspirin and have low potential for NSAID-induced toxicity, should be considered for a non-NSAID agent or use of a traditional NSAID such as naproxen (lower CV side effects) in combination with gastric protection, if warranted. Finally, individuals with CV and GI risks who require aspirin must be considered for non-NSAID therapy, but if that is not an option, then gastric protection with any type of NSAID must be considered. Any patient, regardless of risk status, who is being considered for long-term traditional NSAID therapy should also be considered for *H. pylori* testing and treatment if positive. Assuring the use of GI protective agents with NSAIDs is difficult, even in high-risk patients. This is in part due to underprescribing of the appropriate protective agent; other times, the difficulty is related to patient compliance. The latter may be due to patients forgetting to take multiple pills or preferring not to take the extra pill, especially if they have no GI symptoms. Several NSAID gastroprotective-containing combination pills are now commercially available, including double-dose famotidine with ibuprofen, diclofenac with misoprostol, and naproxen with esomeprazole. Although initial studies suggested improved compliance and a cost advantage when taking these combination drugs, their clinical benefit over the use of separate pills has not been established. One additional concern with NSAID-induced GI complications is the relatively low rate of primary care provider compliance with established guidelines outlining preventative measures. An intervention including professional education, informatics to facilitate review, and financial incentives for practices to review patients' charts to assess appropriateness showed a reduced rate of high-risk prescribing of antiplatelet medications and NSAIDs with a tendency toward improved clinical outcomes. Efforts continue toward developing safer NSAIDs, including topical NSAIDs, NSAID formulations that are rapidly absorbed (diclofenac potassium powder mixed with a buffering agent, Prosorb and SoluMatrix technology), NO-releasing NSAIDs, hydrogen sulfide-releasing NSAIDs, dual COX/5-LOX inhibitors, NSAID prodrugs, and agents that can effectively sequester unbound NSAIDs without interfering with their efficacy.

APPROACH AND THERAPY: SUMMARY

Controversy continues regarding the best approach to the patient who presents with dyspepsia (Chap. 45). The discovery of *H. pylori* and its role in pathogenesis of ulcers has added a new variable to the equation. Previously, if a patient <50 years of age presented with dyspepsia and without alarming signs or symptoms suggestive of an ulcer complication or malignancy, an empirical therapeutic trial with acid suppression was commonly recommended. Although this approach is practiced by some today, an approach presently gaining approval for the treatment of patients with dyspepsia is outlined in Fig. 324-14. The referral to a gastroenterologist is for the potential need of endoscopy and subsequent evaluation and treatment if the endoscopy is negative.

Once an ulcer (GU or DU) is documented, the main issue at stake is whether *H. pylori* or an NSAID is involved. With *H. pylori*

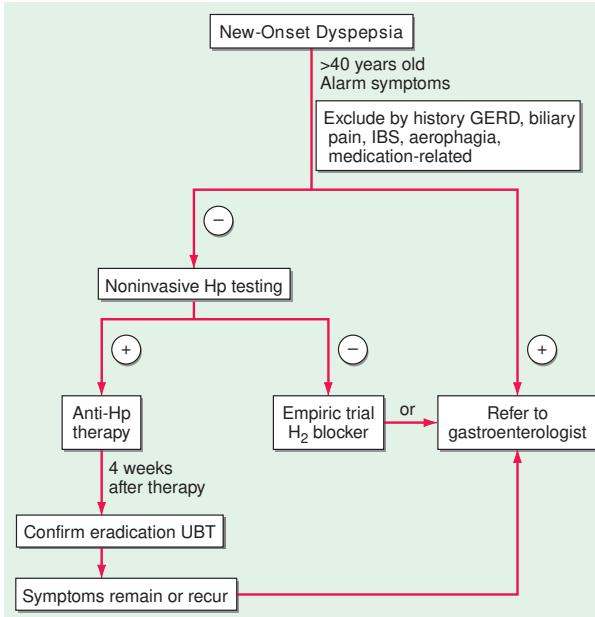


FIGURE 324-14 Overview of new-onset dyspepsia. GERD, gastroesophageal reflux disease; Hp, *Helicobacter pylori*; IBS, irritable bowel syndrome; UBT, urea breath test. (Reproduced with permission from BS Anand, DY Graham: *State-of-the-Art: Ulcer and Gastritis*, Endoscopy 31:215, 1999. © Georg Thieme Verlag KG.)

present, independent of the NSAID status, triple therapy is recommended for 14 days, followed by continued acid-suppressing drugs (H_2 receptor antagonist or PPIs) for a total of 4–6 weeks. *H. pylori* eradication should be documented 4 weeks after completing antibiotics. The test of choice for documenting eradication is the laboratory-based validated monoclonal stool antigen test or a urea breath test (UBT). The patient must be off antisecretory agents for at least 7 days when being tested for eradication of *H. pylori* with UBT or stool antigen. Serologic testing is not useful for the purpose of documenting eradication because antibody titers fall slowly and often do not become undetectable. Some recommend that patients with complicated ulcer disease or who are frail should be treated with long-term acid suppression, thus making documentation of *H. pylori* eradication a moot point. In view of this discrepancy in practice, it would be best to discuss with the patient the different options available.

Several issues differentiate the approach to a GU versus a DU. GUs, especially of the body and fundus, have the potential of being malignant. Multiple biopsies of a GU should be taken initially; even if these are negative for neoplasm, repeat endoscopy to document healing at 8–12 weeks should be performed, with biopsy if the ulcer is still present. About 70% of GUs eventually found to be malignant undergo significant (usually incomplete) healing. Repeat endoscopy is warranted in patients with DU if symptoms persist despite medical therapy or a complication is suspected.

The majority (>90%) of GUs and DUs heal with the conventional therapy outlined above. A GU that fails to heal after 12 weeks and a DU that does not heal after 8 weeks of therapy should be considered refractory. Once poor compliance and persistent *H. pylori* infection have been excluded, NSAID use, either inadvertent or surreptitious, must be excluded. In addition, cigarette smoking must be eliminated. For a GU, malignancy must be meticulously excluded. Next, consideration should be given to a gastric acid hypersecretory state such as ZES (see “Zollinger-Ellison Syndrome,” below) or the idiopathic form, which can be excluded with gastric acid analysis. Although a subset of patients has gastric acid hypersecretion of unclear etiology as a contributing factor to refractory ulcers, ZES should be excluded with a fasting gastrin or secretin stimulation test (see below). More than 90% of refractory ulcers (either DUs

or GUs) heal after 8 weeks of treatment with higher doses of PPI (omeprazole 40 mg/d; lansoprazole 30–60 mg/d). This higher dose is also effective in maintaining remission. Surgical intervention may be a consideration at this point; however, other rare causes of refractory ulcers must be excluded before recommending surgery. Rare etiologies of refractory ulcers that may be diagnosed by gastric or duodenal biopsies include ischemia, Crohn’s disease, amyloidosis, sarcoidosis, lymphoma, eosinophilic gastroenteritis, smoking crack cocaine, or infection (cytomegalovirus [CMV], tuberculosis, or syphilis).

SURGICAL THERAPY

Surgical intervention in PUD can be viewed as being either elective, for treatment of medically refractory disease, or as urgent/emergent, for the treatment of an ulcer-related complication. The development of pharmacologic and endoscopic approaches for the treatment of peptic disease and its complications has led to a substantial decrease in the number of operations needed for this disorder with a drop of >90% for elective ulcer surgery over the past four decades. Refractory ulcers are an exceedingly rare occurrence. Surgery is more often required for treatment of an ulcer-related complication.

Hemorrhage is the most common ulcer-related complication, occurring in ~15–25% of patients. Bleeding may occur in any age group but is most often seen in older patients (sixth decade or beyond). The majority of patients stop bleeding spontaneously, but endoscopic therapy (Chap. 322) is necessary in some. Parenterally and orally administered PPIs also decrease ulcer rebleeding in patients who have undergone endoscopic therapy. Patients unresponsive or refractory to endoscopic intervention will require angiographic intervention or surgery (~5% of transfusion-requiring patients).

Free peritoneal perforation occurs in ~2–3% of DU patients, with NSAID-induced GU perforations occurring more commonly. Sudden onset of severe abdominal pain with peritoneal signs and evidence of pneumoperitoneum on abdominal imaging is the classic presentation of a perforated viscous, but this presentation occurs in only two-thirds of patients. The latter is especially true in elderly patients (>70 years old), obese individuals, and immunocompromised patients. It is important to keep in mind that, as in the case of bleeding, up to 10% of these patients will not have antecedent ulcer symptoms. Delay in diagnosis clearly leads to higher mortality; thus, early suspicion and intervention with nasogastric suction, intravenous PPI, antibiotics and surgical consultation are essential. Concomitant bleeding may occur in up to 10% of patients with perforation, with mortality being increased substantially. Peptic ulcer can also penetrate into adjacent organs, especially with a posterior DU, which can penetrate into the pancreas, colon, liver, or biliary tree.

Pyloric channel ulcers or DUs can lead to gastric outlet obstruction in ~2–3% of patients. This can result from chronic scarring or from impaired motility due to inflammation and/or edema with pylorospasm. Patients may present with early satiety, nausea, vomiting of undigested food, and weight loss. Conservative management with nasogastric suction, intravenous hydration/nutrition, and antisecretory agents is indicated for 7–10 days with the hope that a functional obstruction will reverse. If a mechanical obstruction persists, endoscopic intervention with balloon dilation may be effective. Surgery should be considered if all else fails.

Specific Operations for Duodenal Ulcers Surgical treatment was originally designed to decrease gastric acid secretion. Operations most commonly performed include (1) vagotomy and drainage (by pyloroplasty, gastroduodenostomy, or gastrojejunostomy), (2) highly selective vagotomy (which does not require a drainage procedure), and (3) vagotomy with antrectomy. The specific procedure performed is dictated by the underlying circumstances: elective versus emergency, the degree and extent of duodenal ulceration, the etiology of the ulcer (*H. pylori*, NSAIDs, malignancy), and the expertise of the surgeon. Moreover, the trend has been toward a

dramatic decrease in the need for surgery for treatment of refractory PUD, and when needed, minimally invasive and anatomy-preserving operations are preferred.

Vagotomy is a component of each of these procedures and is aimed at decreasing acid secretion through ablating cholinergic input to the stomach. Unfortunately, both truncal and selective vagotomy (preserves the celiac and hepatic branches) result in gastric atony despite successful reduction of both basal acid output (BAO; decreased by 85%) and maximal acid output (MAO; decreased by 50%). Drainage through pyloroplasty or gastroduodenostomy is required in an effort to compensate for the vagotomy-induced gastric motility disorder. This procedure has an intermediate complication rate and a 10% ulcer recurrence rate. To minimize gastric dysmotility, highly selective vagotomy (also known as parietal cell, super-selective, or proximal vagotomy) was developed. Only the vagal fibers innervating the portion of the stomach that contains parietal cells are transected, thus leaving fibers important for regulating gastric motility intact. Although this procedure leads to an immediate decrease in both BAO and stimulated acid output, acid secretion recovers over time. By the end of the first postoperative year, basal and stimulated acid output are ~30 and 50%, respectively, of preoperative levels. Ulcer recurrence rates are higher with highly selective vagotomy ($\geq 10\%$), although the overall complication rates are the lowest of the three procedures.

The procedure that provides the lowest rates of ulcer recurrence (1%) but has the highest complication rate is vagotomy (truncal or selective) in combination with antrectomy. Antrectomy is aimed at eliminating an additional stimulant of gastric acid secretion, gastrin. Two principal types of reanastomoses are used after antrectomy: gastroduodenostomy (Billroth I) or gastrojejunostomy (Billroth II) (Fig. 324-15). Although Billroth I is often preferred over II, severe duodenal inflammation or scarring may preclude its performance. Prospective, randomized studies confirm that partial gastrectomy followed by Roux-en-Y reconstruction leads to a significantly better clinical, endoscopic, and histologic outcome than Billroth II reconstruction.

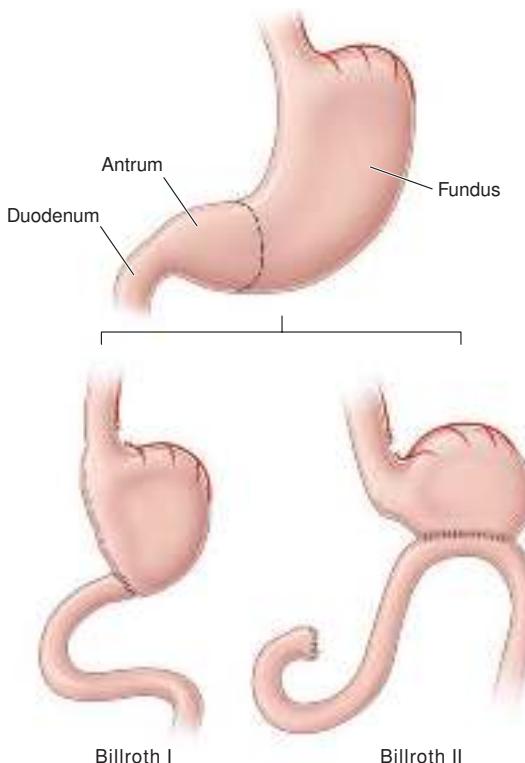


FIGURE 324-15 Schematic representation of Billroth I and II procedures.

Of these procedures, highly selective vagotomy may be the procedure of choice in the elective setting, except in situations where ulcer recurrence rates are high (prepyloric ulcers and those refractory to medical therapy). Selection of vagotomy and antrectomy may be more appropriate in these circumstances.

These procedures have been traditionally performed by standard laparotomy. The advent of laparoscopic surgery has led several surgical teams to successfully perform highly selective vagotomy, truncal vagotomy/pyloroplasty, and truncal vagotomy/antrectomy through this approach. An increase in the number of laparoscopic procedures for treatment of PUD has occurred. Laparoscopic repair of perforated peptic ulcers is safe, feasible for the experienced surgeon, and associated with decreased postoperative pain, although it does take longer than an open approach. Moreover, no difference between the two approaches is noted in postoperative complications or length of hospital stay.

Specific Operations for GUs The location and presence of a concomitant DU dictate the operative procedure performed for a GU. Antrectomy (including the ulcer) with a Billroth I anastomosis is the treatment of choice for an antral ulcer. Vagotomy is performed only if a DU is present. Although ulcer excision with vagotomy and drainage procedure has been proposed, the higher incidence of ulcer recurrence makes this a less desirable approach. Ulcers located near the esophagogastric junction may require a more radical approach, a subtotal gastrectomy with a Roux-en-Y esophagogastrojejunostomy (Csendes' procedure). A less aggressive approach, including antrectomy, intraoperative ulcer biopsy, and vagotomy (Kelling-Madlener procedure), may be indicated in fragile patients with a high GU. Ulcer recurrence approaches 30% with this procedure.

Surgery-Related Complications Complications seen after surgery for PUD are related primarily to the extent of the anatomic modification performed. Minimal alteration (highly selective vagotomy) is associated with higher rates of ulcer recurrence and less GI disturbance. More aggressive surgical procedures have a lower rate of ulcer recurrence but a greater incidence of GI dysfunction. Overall, morbidity and mortality related to these procedures are quite low. Morbidity associated with vagotomy and antrectomy or pyloroplasty is $\leq 5\%$, with mortality $\sim 1\%$. Highly selective vagotomy has lower morbidity and mortality rates of 1 and 0.3%, respectively.

In addition to the potential early consequences of any intraabdominal procedure (bleeding, infection, thromboembolism), gastroparesis, duodenal stump leak, and efferent loop obstruction can be observed.

Recurrent Ulceration The risk of ulcer recurrence is directly related to the procedure performed. Ulcers that recur after partial gastric resection tend to develop at the anastomosis (stomal or marginal ulcer). Epigastric abdominal pain is the most frequent presenting complaint (>90%). Severity and duration of pain tend to be more progressive than observed with DUs before surgery.

Ulcers may recur for several reasons, including incomplete vagotomy, inadequate drainage, retained antrum, and, less likely, persistent or recurrent *H. pylori* infection. ZES should have been excluded preoperatively. Surreptitious use of NSAIDs is an important reason for recurrent ulcers after surgery, especially if the initial procedure was done for an NSAID-induced ulcer. Once *H. pylori* and NSAIDs have been excluded as etiologic factors, the question of incomplete vagotomy or retained gastric antrum should be explored. For the latter, fasting plasma gastrin levels should be determined. If elevated, retained antrum or ZES (see below) should be considered. Incomplete vagotomy can be ruled out by gastric acid analysis coupled with sham feeding. In this test, gastric acid output is measured while the patient sees, smells, and chews a meal (without swallowing). The cephalic phase of gastric secretion, which is mediated by the vagus, is being assessed with this study. An increase in gastric acid output in response to sham feeding is evidence that the vagus nerve is intact. A rise in serum pancreatic polypeptide >50% within 30 min of sham feeding is also suggestive of an intact vagus nerve.

Medical therapy with H₂ blockers will heal postoperative ulceration in 70–90% of patients. The efficacy of PPIs has not been fully assessed in this group, but one may anticipate greater rates of ulcer healing compared to those obtained with H₂ blockers. Repeat operation (complete vagotomy, partial gastrectomy) may be required in a small subgroup of patients who have not responded to aggressive medical management.

Afferent Loop Syndromes Although rarely seen today as a result of the decrease in the performance of Billroth II anastomosis, two types of afferent loop syndrome can occur in patients who have undergone this type of partial gastric resection. The more common of the two is bacterial overgrowth in the afferent limb secondary to stasis. Patients may experience postprandial abdominal pain, bloating, and diarrhea with concomitant malabsorption of fats and vitamin B₁₂. Cases refractory to antibiotics may require surgical revision of the loop. The less common afferent loop syndrome can present with severe abdominal pain and bloating that occur 20–60 min after meals. Pain is often followed by nausea and vomiting of bile-containing material. The pain and bloating may improve after emesis. The cause of this clinical picture is theorized to be incomplete drainage of bile and pancreatic secretions from an afferent loop that is partially obstructed. Cases refractory to dietary measures may need surgical revision or conversion of the Billroth II anastomosis to a Roux-en-Y gastrojejunostomy.

Dumping Syndrome Dumping syndrome consists of a series of vaso-motor and GI signs and symptoms and occurs in patients who have undergone vagotomy and drainage (especially Billroth procedures). Two phases of dumping, early and late, can occur. Early dumping takes place 15–30 min after meals and consists of crampy abdominal discomfort, nausea, diarrhea, belching, tachycardia, palpitations, diaphoresis, light-headedness, and, rarely, syncope. These signs and symptoms arise from the rapid emptying of hyperosmolar gastric contents into the small intestine, resulting in a fluid shift into the gut lumen with plasma volume contraction and acute intestinal distention. Release of vasoactive GI hormones (vasoactive intestinal polypeptide, neurotensin, motilin) is also theorized to play a role in early dumping.

The late phase of dumping typically occurs 90 min to 3 h after meals. Vasomotor symptoms (light-headedness, diaphoresis, palpitations, tachycardia, and syncope) predominate during this phase. This component of dumping is thought to be secondary to hypoglycemia from excessive insulin release.

Dumping syndrome is most noticeable after meals rich in simple carbohydrates (especially sucrose) and high osmolarity. Ingestion of large amounts of fluids may also contribute. After vagotomy and drainage, up to 50% of patients will experience dumping syndrome to some degree early on. Signs and symptoms often improve with time, but a severe protracted picture can occur in up to 1% of patients.

Dietary modification is the cornerstone of therapy for patients with dumping syndrome. Small, multiple (six) meals devoid of simple carbohydrates coupled with elimination of liquids during meals is important. Antidiarrheals and anticholinergic agents are complementary to diet. Guar and pectin, which increase the viscosity of intraluminal contents, may be beneficial in more symptomatic individuals. Acarbose, an α-glucosidase inhibitor that delays digestion of ingested carbohydrates, has also been shown to be beneficial in the treatment of the late phases of dumping. The somatostatin analogue octreotide has been successful in diet-refractory cases. This drug is administered subcutaneously (50 µg tid), titrated according to clinical response. A long-acting depot formulation of octreotide can be administered once every 28 days and provides symptom relief comparable to the short-acting agent. In addition, patient weight gain and quality of life appear to be superior with the long-acting form.

Postvagotomy Diarrhea Up to 10% of patients may seek medical attention for the treatment of postvagotomy diarrhea. This complication is most commonly observed after truncal vagotomy, which is rarely performed today. Patients may complain of intermittent

diarrhea that occurs typically 1–2 h after meals. Occasionally, the symptoms may be severe and relentless. This is due to a motility disorder from interruption of the vagal fibers supplying the luminal gut. Other contributing factors may include decreased absorption of nutrients (see below), increased excretion of bile acids, and release of luminal factors that promote secretion. Diphenoxylate or loperamide is often useful in symptom control. The bile salt-binding agent cholestyramine may be helpful in severe cases. Surgical reversal of a 10-cm segment of jejunum may yield a substantial improvement in bowel frequency in a subset of patients.

Bile Reflux Gastropathy A subset of post-partial gastrectomy patients who present with abdominal pain, early satiety, nausea, and vomiting will have mucosal erythema of the gastric remnant as the only finding. Histologic examination of the gastric mucosa reveals minimal inflammation but the presence of epithelial cell injury. This clinical picture is categorized as bile or alkaline reflux gastropathy/gastritis. Although reflux of bile is implicated as the reason for this disorder, the mechanism is unknown. Prokinetic agents, cholestyramine, and sucralfate have been somewhat effective treatments. Severe refractory symptoms may require using either nuclear scanning with ^{99m}Tc-HIDA to document reflux. Surgical diversion of pancreaticobiliary secretions away from the gastric remnant with a Roux-en-Y gastrojejunostomy consisting of a long (50–60 cm) Roux limb has been used in severe cases. Bilious vomiting improves, but early satiety and bloating may persist in up to 50% of patients.

Maldigestion and Malabsorption Weight loss can be observed in up to 60% of patients after partial gastric resection. Patients can experience a 10% loss of body weight, which stabilizes 3 months postoperatively. A significant component of this weight reduction is due to decreased oral intake. However, mild steatorrhea can also develop. Reasons for maldigestion/malabsorption include decreased gastric acid production, rapid gastric emptying, decreased food dispersion in the stomach, reduced luminal bile concentration, reduced pancreatic secretory response to feeding, and rapid intestinal transit.

Decreased serum vitamin B₁₂ levels can be observed after partial gastrectomy. This is usually not due to deficiency of IF, since a minimal amount of parietal cells (source of IF) is removed during antrectomy. Reduced vitamin B₁₂ may be due to competition for the vitamin by bacterial overgrowth or inability to split the vitamin from its protein-bound source due to hypochlorhydria.

Iron-deficiency anemia may be a consequence of impaired absorption of dietary iron in patients with a Billroth II gastrojejunostomy. Absorption of iron salts is normal in these individuals; thus, a favorable response to oral iron supplementation can be anticipated. Folate deficiency with concomitant anemia can also develop in these patients. This deficiency may be secondary to decreased absorption or diminished oral intake.

Malabsorption of vitamin D and calcium resulting in osteoporosis and osteomalacia is common after partial gastrectomy and gastrojejunostomy (Billroth II). Osteomalacia can occur as a late complication in up to 25% of post-partial gastrectomy patients. Bone fractures occur twice as commonly in men after gastric surgery as in a control population. It may take years before x-ray findings demonstrate diminished bone density. Elevated alkaline phosphatase, reduced serum calcium, bone pain, and pathologic fractures may be seen in patients with osteomalacia. The high incidence of these abnormalities in this subgroup of patients justifies treating them with vitamin D and calcium supplementation indefinitely. Therapy is especially important in females. Copper deficiency has also been reported in patients undergoing surgeries that bypass the duodenum, where copper is primarily absorbed. Patients may present with a rare syndrome that includes ataxia, myelopathy, and peripheral neuropathy.

Gastric Adenocarcinoma The incidence of adenocarcinoma in the gastric stump is increased 15 years after resection. Some have reported a four- to fivefold increase in gastric cancer 20–25 years

after resection. The pathogenesis is unclear but may involve alkaline reflux, bacterial proliferation, or hypochlorhydria. The role of endoscopic screening is not clear, and most guidelines do not support its use.

Additional Complications Reflux esophagitis and a higher incidence of gallstones and cholecystitis have been reported in patients undergoing subtotal gastrectomy. The latter is thought to be due to decreased gallbladder contractility associated with vagotomy and bypass of the duodenum, leading to decreased postprandial release of cholecystokinin.

RELATED CONDITIONS

ZOLLINGER ELLISON SYNDROME

Severe peptic ulcer diathesis secondary to gastric acid hypersecretion due to unregulated gastrin release from a non- β -cell, often well-differentiated neuroendocrine tumor (NET; gastrinoma) defines the components of ZES. Initially, ZES was typified by aggressive and refractory ulceration in which total gastrectomy provided the only chance for enhancing survival. Today, it can be cured by surgical resection in up to 40% of patients with the sporadic form of the disease (see below).

Epidemiology The true incidence of ZES is unknown, but estimates suggest that it varies from 0.1 to 1% of individuals presenting with PUD, with 0.1–3 individuals per year having this rare diagnosis. Others have estimated an incidence of 0.5–3 per million population. Females are slightly more commonly affected than males, and the majority of patients are diagnosed between ages 30 and 50. Gastrinomas are classified into sporadic tumors (80%) and those associated with multiple endocrine neoplasia (MEN) type 1 (see below). The widespread availability and use of PPIs have led to a decreased patient referral for gastrinoma evaluation, delay in diagnosis, and an increase in false-positive diagnoses of ZES. In fact, diagnosis may be delayed for ≥ years after symptoms consistent with ZES are displayed.

Pathophysiology Hypergastrinemia originating from an autonomous neoplasm is the driving force responsible for the clinical manifestations in ZES. Gastrin stimulates acid secretion through gastrin receptors on parietal cells and by inducing histamine release from ECL cells. Gastrin also has a trophic action on gastric epithelial cells. Long-standing hypergastrinemia leads to markedly increased gastric acid secretion through both parietal cell stimulation and increased parietal cell mass. The increased gastric acid output leads to peptic ulcer diathesis, erosive esophagitis, and diarrhea.

Tumor Distribution Although early studies suggested that the vast majority of gastrinomas occurred within the pancreas, a significant number of these lesions are extrapancreatic. Between 60 and 90% of these tumors are found within the hypothetical gastrinoma triangle (confluence of the cystic and common bile ducts superiorly, junction of the second and third portions of the duodenum inferiorly, and junction of the neck and body of the pancreas medially). Duodenal tumors constitute the most common nonpancreatic lesion; between 60 and 100% of gastrinomas are found here. Duodenal tumors are smaller, slower growing, and less likely to metastasize than pancreatic lesions. Less common extrapancreatic sites include stomach, bones, ovaries, heart, liver, and lymph nodes. More than 60% of tumors are considered malignant, with up to 30–50% of patients having multiple lesions or metastatic disease at presentation. Histologically, gastrin-producing cells appear well-differentiated (grade 1 or 2 histologically), expressing markers typically found in endocrine neoplasms (chromogranin, neuron-specific enolase). Although not clearly established in gastrinomas, histologic grade in pancreatic NETs generally is an important predictor of survival in these rare neoplasms (Chap. 84).

Clinical Manifestations Gastric acid hypersecretion is responsible for the signs and symptoms observed in patients with ZES. The most common clinical presentation for gastrinoma patients is abdominal pain in the presence of acid peptic disorders. Peptic ulcer is the

most common clinical manifestation, occurring in >90% of gastrinoma patients. Initial presentation and ulcer location (duodenal bulb) may be indistinguishable from common PUD. Clinical situations that should create suspicion of gastrinoma are ulcers in unusual locations (second part of the duodenum and beyond), ulcers refractory to standard medical therapy, ulcer recurrence after acid-reducing surgery, ulcers presenting with frank complications (bleeding, obstruction, and perforation), or ulcers in the absence of *H. pylori* or NSAID ingestion. Symptoms of esophageal origin are present in up to two-thirds of patients with ZES, with a spectrum ranging from mild esophagitis to frank ulceration with stricture and Barrett's mucosa.

Diarrhea, the next most common clinical manifestation, is found in up to 70% of patients. Although diarrhea often occurs concomitantly with acid peptic disease, it may also occur independent of an ulcer and classically will abate with PPI therapy. Etiology of the diarrhea is multifactorial, resulting from marked volume overload to the small bowel, pancreatic enzyme inactivation by acid, and damage of the intestinal epithelial surface by acid. The epithelial damage can lead to a mild degree of maldigestion and malabsorption of nutrients. The diarrhea may also have a secretory component due to the direct stimulatory effect of gastrin on enterocytes or the co-secretion of additional hormones from the tumor such as vasoactive intestinal peptide.

Gastrinomas can develop in the presence of MEN 1 syndrome (Chaps. 84 and 388) in ~25% of patients. This autosomal dominant disorder involves primarily three organ sites: the parathyroid glands (80–90%), pancreas (40–80%), and pituitary gland (30–60%). The syndrome is caused by inactivating mutations of the *MEN1* tumor-suppressor gene found on the long arm of chromosome 11q13. The gene encodes for menin, which has an important role in DNA replication and transcriptional regulation. A genetic diagnosis is obtained by sequencing of the *MEN1* gene, which can reveal mutations in 70–90% of typical MEN 1 cases. A family may have an unknown mutation, making a genetic diagnosis impossible, and therefore, certain individuals will require a clinical diagnosis, which is determined by whether a patient has tumors in two of the three endocrine organs (parathyroid, pancreas/duodenum, or pituitary) or has a family history of MEN 1 and one of the endocrine organ tumors. In view of the stimulatory effect of calcium on gastric secretion, the hyperparathyroidism and hypercalcemia seen in MEN 1 patients may have a direct effect on ulcer disease. Resolution of hypercalcemia by parathyroidectomy reduces gastrin and gastric acid output in gastrinoma patients. An additional distinguishing feature in ZES patients with MEN 1 is the higher incidence of gastric carcinoid tumor development (as compared to patients with sporadic gastrinomas). ZES presents and is diagnosed earlier in MEN 1 patients, and they have a more indolent course as compared to patients with sporadic gastrinoma. Gastrinomas tend to be smaller, multiple, and located in the duodenal wall more often than is seen in patients with sporadic ZES. Establishing the diagnosis of MEN 1 is critical in order to provide genetic counseling to the patient and his or her family and also to determine the recommended surgical approach. Therefore, gastrinoma patients should be screened for MEN 1 by performing a detailed family history and obtaining several serum markers including calcium, parathyroid, prolactin, and pancreatic polypeptide levels.

Diagnosis Establishing an early diagnosis is important in order to minimize the long-term sequelae of gastric acid hypersecretion, prevent metastatic disease, and counsel family members if a diagnosis of MEN 1 is established. Biochemical measurements of gastrin and acid secretion in patients suspected of having ZES play an important role in establishing this rare diagnosis. Often, patients suspected of having ZES will be treated with a PPI in an effort to ameliorate symptoms and decrease the likelihood of possible acid-related complications. The presence of the PPI, which will lower acid secretion and potentially elevate fasting gastrin levels in normal individuals, will make the diagnostic approach in these individuals somewhat difficult. Significant morbidity related to peptic diathesis has been described when stopping PPIs in gastrinoma patients; therefore, a systematic approach in stopping these agents is warranted (see below). The first step in the

TABLE 324-8 When to Obtain a Fasting Serum Gastrin Level

Multiple ulcers
Ulcers in unusual locations; associated with severe esophagitis; resistant to therapy with frequent recurrences; in the absence of nonsteroidal anti-inflammatory drug ingestion or <i>H. pylori</i> infection
Ulcer patients awaiting surgery
Extensive family history for peptic ulcer disease
Postoperative ulcer recurrence
Basal hyperchlorhydria
Unexplained diarrhea or steatorrhea
Hypercalcemia
Family history of pancreatic islet, pituitary, or parathyroid tumor
Prominent gastric or duodenal folds

evaluation of a patient suspected of having ZES is to obtain a fasting gastrin level. A list of clinical scenarios that should arouse suspicion regarding this diagnosis is shown in Table 324-8. Fasting gastrin levels obtained using a dependable assay are usually <150 pg/mL. A normal fasting gastrin, on two separate occasions, especially if the patient is on a PPI, virtually excludes this diagnosis. Virtually all gastrinoma patients will have a gastrin level >150–200 pg/mL. Measurement of fasting gastrin should be repeated to confirm the clinical suspicion. Some of the commercial biochemical assays used for measuring serum gastrin may be inaccurate. Variable specificity of the antibodies used have led to both false-positive and false-negative fasting gastrin levels, placing in jeopardy the ability to make an accurate diagnosis of ZES.

Multiple processes can lead to an elevated fasting gastrin level, the most frequent of which are gastric hypochlorhydria and achlorhydria, with or without pernicious anemia. Gastric acid induces feedback inhibition of gastrin release. A decrease in acid production will subsequently lead to failure of the feedback inhibitory pathway, resulting in net hypergastrinemia. Gastrin levels will thus be high in patients using antisecretory agents for the treatment of acid peptic disorders and dyspepsia. *H. pylori* infection can also cause hypergastrinemia. Additional causes of elevated gastrin include retained gastric antrum; G-cell hyperplasia; gastric outlet obstruction; renal insufficiency; massive small-bowel obstruction; and conditions such as rheumatoid arthritis, vitiligo, diabetes mellitus, and pheochromocytoma. Although a fasting gastrin >10 times normal is highly suggestive of ZES, two-thirds of patients will have fasting gastrin levels that overlap with levels found in the more common disorders outlined above, especially if a PPI is being taken by the patient. The effect of the PPI on gastrin levels and acid secretion will linger several days after stopping the PPI; therefore, it should be stopped for a minimum of 7 days before testing. During this period, the patient should be placed on a histamine H₂ antagonist, such as famotidine, twice to three times per day. Although this type of agent has a short-term effect on gastrin and acid secretion, it needs to be stopped 24 h before repeating fasting gastrin levels or performing some of the tests highlighted below. The patient may take antacids for the final day, stopping them ~12 h before testing is performed. Heightened awareness of complications related to gastric acid hypersecretion during the period of PPI cessation is critical.

The next step at times needed for establishing a biochemical diagnosis of gastrinoma is to assess acid secretion. Nothing further needs to be done if decreased acid output in the absence of a PPI is observed. A pH can be measured on gastric fluid obtained either during endoscopy or through nasogastric aspiration; a pH <3 is suggestive of a gastrinoma, but a pH >3 is not helpful in excluding the diagnosis. In those situations where the pH is >3, formal gastric acid analysis should be performed if available. Normal BAO in nongastric surgery patients is typically <5 meq/h. A BAO >15 meq/h in the presence of hypergastrinemia is considered pathognomonic of ZES, but up to 12% of patients with common PUD may have elevated BAO to a lesser degree that can overlap with levels seen in ZES patients. In an effort to improve the sensitivity and specificity of gastric secretory studies, a BAO/MAO ratio was established using pentagastrin infusion as a way to maximally stimulate acid production, with a BAO/MAO ratio >0.6

being highly suggestive of ZES. Pentagastrin is no longer available in the United States, making measurement of MAO virtually impossible. An endoscopic method for measuring gastric acid output has been developed but requires further validation.

Gastrin provocative tests have been developed in an effort to differentiate between the causes of hypergastrinemia and are especially helpful in patients with indeterminate acid secretory studies. The tests are the secretin stimulation test and the calcium infusion study; the latter is rarely, if ever, utilized in our current environment due to the cumbersome nature of the test and its lower sensitivity and specificity than secretin stimulation. The most sensitive and specific gastrin provocative test for the diagnosis of gastrinoma is the secretin study. An increase in gastrin of ≥120 pg within 15 min of secretin injection has a sensitivity and specificity of >90% for ZES. PPI-induced hypochlorhydria or achlorhydria may lead to a false-positive secretin test; thus, this agent must be stopped for 1 week before testing.

In light of the limited availability of the biochemical studies outlined above, more studies make a diagnosis of gastrinoma based on the presence of elevated gastrin and low gastric pH in the right clinical setting coupled with tumor localization tests outlined below and positive histology by biopsy (difficult to obtain). Revised guidelines for the best approach to establishing a diagnosis of gastrinoma taking into consideration the above outlined limitations are being considered, but none have replaced the established guidelines outlined earlier in this section.

Tumor Localization Once the biochemical diagnosis of gastrinoma has been confirmed (if possible), the tumor must be located. Multiple imaging studies have been used in an effort to enhance tumor localization (Table 324-9). The broad range of sensitivity is due to the variable success rates achieved by the different investigative groups. Endoscopic ultrasound (EUS) permits imaging of the pancreas with high degree of resolution (<5 mm). This modality is particularly helpful in excluding small neoplasms within the pancreas and in assessing the presence of surrounding lymph nodes and vascular involvement, but it is not very sensitive (43%) for finding duodenal lesions. This latter observation has led some to not include EUS in the routine preoperative evaluation of a patient suspected of having a gastrinoma. Several types of endocrine tumors express cell-surface receptors for somatostatin, in particular the subtype 2 (SSTR2). This permits the localization, staging, and prediction of therapeutic response to somatostatin analogues (see below) by gastrinomas. The original functional scintigraphic tool developed measuring the uptake of the stable somatostatin analogue ¹¹¹In-pentetreotide (OctreoScan) has demonstrated sensitivity and specificity rates of >80%. More recently, positron emission tomography (PET)-computed tomography (CT) with ⁶⁸Ga-DOTATATE has been developed and is superior than OctreoScan for assessing tumor presence in patients with well-differentiated NETs such as gastrinomas, with sensitivity and specificity of >90%, making it the functional imaging study of choice when available. ¹⁸F-Fluorodeoxyglucose (¹⁸F-FDG) PET imaging

TABLE 324-9 Sensitivity of Imaging Studies in Zollinger-Ellison Syndrome

STUDY	SENSITIVITY, %	
	PRIMARY GASTRINOMA	METASTATIC GASTRINOMA
Ultrasound	21–28	14
CT scan	55–70	>85
Selective angiography	35–68	33–86
Portal venous sampling	70–90	N/A
SASI	55–78	41
MRI	55–70	>85
OctreoScan	67–86	80–100
EUS	80–100	N/A

Abbreviations: CT, computed tomography; EUS, endoscopic ultrasonography; MRI, magnetic resonance imaging; N/A, not applicable; OctreoScan, imaging with ¹¹¹In-pentetreotide; SASI, selective arterial secretin injection.

has been found to be useful in pancreatic NETs, including gastrinomas, particularly as a prognostic marker.

Up to 50% of patients have metastatic disease at diagnosis. Success in controlling gastric acid hypersecretion has shifted the emphasis of therapy toward providing a surgical cure. Detecting the primary tumor and excluding metastatic disease are critical in view of this paradigm shift. Once a biochemical diagnosis has been confirmed, the patient should first undergo an abdominal CT scan, magnetic resonance imaging (MRI), or OctreoScan/PET-CT with ⁶⁸Ga-DOTATATE (depending on availability) to exclude metastatic disease. Once metastatic disease has been excluded, an experienced endocrine surgeon may opt for exploratory laparotomy with intraoperative ultrasound or transillumination. In other centers, careful examination of the peripancreatic area with EUS, accompanied by endoscopic exploration of the duodenum for primary tumors, will be performed before surgery. Selective arterial secretin injection may be a useful adjuvant for localizing tumors in a subset of patients. The extent of the diagnostic and surgical approach must be carefully balanced with the patient's overall physiologic condition and the natural history of a slow-growing gastrinoma.

TREATMENT

Zollinger-Ellison Syndrome

Treatment of functional endocrine tumors is directed at ameliorating the signs and symptoms related to hormone overproduction, curative resection of the neoplasm, and attempts to control tumor growth in metastatic disease.

PPIs are the treatment of choice and have decreased the need for total gastrectomy. Initial PPI doses tend to be higher than those used for treatment of GERD or PUD. The initial dose of omeprazole, lansoprazole, rabeprazole, or esomeprazole should be in the range of 60 mg in divided doses in a 24-h period. When gastric acid analysis was more widely available, dosing was adjusted to achieve a BAO <10 meq/h (at the drug trough) in surgery-naïve patients and to <5 meq/h in individuals who have previously undergone an acid-reducing operation. Close monitoring of clinical symptoms when starting PPIs and increasing the dose accordingly are paramount. Although the somatostatin analogue has inhibitory effects on gastrin release from receptor-bearing tumors and inhibits gastric acid secretion to some extent, PPIs have the advantage of reducing parietal cell activity to a greater degree. Despite this, octreotide or lanreotide may be considered as adjunctive therapy to the PPI in patients with tumors that express somatostatin receptors and have peptic symptoms that are difficult to control with high-dose PPI.

The ultimate goal of surgery would be to provide a definitive cure. Improved understanding of tumor distribution has led to immediate cure rates as high as 33% with 10-year disease-free intervals as high as 95% in sporadic gastrinoma patients undergoing surgery. A positive outcome is highly dependent on the experience of the surgical team treating these rare tumors. Surgical therapy of gastrinoma patients with MEN 1 remains controversial because of the difficulty in rendering these patients disease-free with surgery. In contrast to the encouraging postoperative results observed in patients with sporadic disease, <5% of MEN 1 patients are disease-free 5 years after an operation. Moreover, in contrast to patients with sporadic ZES, the clinical course of MEN 1 patients tends to be benign and rarely leads to disease-related mortality, recommending that early surgery be deferred. Some groups suggest surgery only if a clearly identifiable, nonmetastatic lesion is documented by structural studies. Others advocate a more aggressive approach, where all patients free of hepatic metastasis are explored and all detected tumors in the duodenum are resected; this is followed by enucleation of lesions in the pancreatic head, with a distal pancreatectomy to follow. The outcome of the two approaches has not been clearly defined. Laparoscopic surgical interventions may provide attractive approaches in the future but currently seem to be of some limited benefit in patients with gastrinoma because a significant percentage of the tumors may be extrapancreatic and difficult to localize with a

laparoscopic approach. Finally, patients selected for surgery should be individuals whose health status would lead them to tolerate a more aggressive operation and obtain the long-term benefits from such aggressive surgery, which are often witnessed after 10 years.

Therapy of metastatic endocrine tumors in general remains suboptimal; gastrinomas are no exception. In light of the observation that in many instances tumor growth is indolent and that many individuals with metastatic disease remain relatively stable for significant periods of time, many advocate not instituting systemic tumor-targeted therapy until evidence of tumor progression or refractory symptoms not controlled with PPIs are noted. Medical approaches, including biologic therapy (IFN- α , long-acting somatostatin analogues, and peptide receptor radionuclides), systemic chemotherapy (streptozotocin, 5-fluorouracil, and doxorubicin), and hepatic artery embolization, may lead to significant toxicity without a substantial improvement in overall survival. Use of temozolamide with capecitabine has demonstrated radiographic regression and progression-free survival in patients with well-differentiated NETs in the range of 70% and 18 months, respectively. Systemic therapy with radiolabeled somatostatin analogues (peptide receptor radiotherapy [PRRT]) has been used in the therapy of metastatic NETs and appears to be very promising in terms of radiographic regression, symptoms, and progression-free survival, but additional studies are warranted. Several promising therapies are being explored, including radiofrequency ablation or cryoablation of liver lesions and use of agents that block the VEGF receptor pathway (sunitinib), the mammalian target of rapamycin, and immune checkpoint inhibitors (Chap. 87).

Surgical approaches, including debulking surgery and liver transplantation for hepatic metastasis, have also produced limited benefit.

The overall 5- and 10-year survival rates for gastrinoma patients are 62–75% and 47–53%, respectively. Individuals with the entire tumor resected or those with a negative laparotomy have 5- and 10-year survival rates >90%. Patients with incompletely resected tumors have 5- and 10-year survival rates of 43 and 25%, respectively. Patients with hepatic metastasis have <20% survival at 5 years. Favorable prognostic indicators include primary duodenal wall tumors, isolated lymph node tumor, the presence of MEN 1, and undetectable tumor upon surgical exploration. Poor outcome is seen in patients with shorter disease duration; female sex; older age at diagnosis; higher gastrin levels (>10,000 pg/mL); poor histologic differentiation; high proliferative index; large pancreatic primary tumors (>3 cm); metastatic disease to lymph nodes, liver, and bone; and Cushing's syndrome. Rapid growth of hepatic metastases is also predictive of poor outcome.

■ STRESS RELATED MUCOSAL INJURY

Patients suffering from shock, sepsis, massive burns, severe trauma, or head injury can develop acute erosive gastric mucosal changes or frank ulceration with bleeding. Classified as stress-induced gastritis or ulcers, injury is most commonly observed in the acid-producing (fundus and body) portions of the stomach. The most common presentation is GI bleeding, which is usually minimal but can occasionally be life-threatening. Respiratory failure requiring mechanical ventilation and underlying coagulopathy are risk factors for bleeding, which tends to occur 48–72 h after the acute injury or insult.

Histologically, stress injury does not contain inflammation or *H. pylori*; thus, "gastritis" is a misnomer. Although elevated gastric acid secretion may be noted in patients with stress ulceration after head trauma (Cushing's ulcer) and severe burns (Curling's ulcer), mucosal ischemia, breakdown of the normal protective barriers of the stomach, systemic release of cytokines, poor GI motility, and oxidative stress also play an important role in the pathogenesis. Acid must contribute to injury in view of the significant drop in bleeding noted when acid inhibitors are used as prophylaxis for stress gastritis.

Improvement in the general management of intensive care unit patients has led to a significant decrease in the incidence of GI bleeding

due to stress ulceration. The estimated decrease in bleeding is from 20–30% to <5%. This improvement has led to some debate regarding the need for prophylactic therapy. The high mortality associated with stress-induced clinically important GI bleeding (>40%) and the limited benefit of medical (endoscopic, angiographic) and surgical therapy in a patient with hemodynamically compromising bleeding associated with stress ulcer/gastritis support the use of preventive measures in high-risk patients (mechanically ventilated, coagulopathy, multiorgan failure, or severe burns). Meta-analysis comparing H₂ blockers with PPIs for the prevention of stress-associated clinically important and overt GI bleeding demonstrates superiority of the latter without increasing the risk of nosocomial infections, increasing mortality, or prolonging intensive care unit length of stay. Therefore, PPIs are the treatment of choice for stress prophylaxis. Oral PPI is the best option if the patient can tolerate enteral administration. Pantoprazole is available as an intravenous formulation for individuals in whom enteral administration is not possible. If bleeding occurs despite these measures, endoscopy, intraarterial vasopressin, and embolization are options. If all else fails, then surgery should be considered. Although vagotomy and antrectomy may be used, the better approach would be a total gastrectomy, which has an exceedingly high mortality rate in this setting. Concerns with the effect of PPIs on the immune system coupled with the high cost of this agent have led to several comparative studies of PPIs and H₂ receptor antagonists for stress prophylaxis in patients requiring mechanical ventilation. Although the PEPTIC trial demonstrated comparative efficacy between the two agents regarding mortality, technical aspects of the study led to some limitation in the final interpretation of the results.

GASTRITIS

The term *gastritis* should be reserved for histologically documented inflammation of the gastric mucosa. Gastritis is not the mucosal erythema seen during endoscopy and is not interchangeable with “dyspepsia.” The etiologic factors leading to gastritis are broad and heterogeneous. Gastritis has been classified based on time course (acute vs chronic), histologic features, and anatomic distribution or proposed pathogenic mechanism (Table 324-10).

The correlation between the histologic findings of gastritis, the clinical picture of abdominal pain or dyspepsia, and endoscopic findings noted on gross inspection of the gastric mucosa is poor. Therefore, there is no typical clinical manifestation of gastritis.

Acute Gastritis The most common causes of acute gastritis are infectious. Acute infection with *H. pylori* induces gastritis. However,

TABLE 324-10 Classification of Gastritis

- I. Acute gastritis
 - A. Acute *Helicobacter pylori* infection
 - B. Other acute infectious gastritides
 - 1. Bacterial (other than *H. pylori*)
 - 2. *Helicobacter heilmannii*
 - 3. Phlegmonous
 - 4. Mycobacterial
 - 5. Syphilitic
 - 6. Viral
 - 7. Parasitic
 - 8. Fungal
- II. Chronic atrophic gastritis
 - A. Type A: Autoimmune, body-predominant
 - B. Type B: *H. pylori*-related, antral-predominant
 - C. Indeterminate
- III. Uncommon forms of gastritis
 - A. Lymphocytic
 - B. Eosinophilic
 - C. Crohn's disease
 - D. Sarcoidosis
 - E. Isolated granulomatous gastritis
 - F. Russell body gastritis

H. pylori acute gastritis has not been extensively studied. It is reported as presenting with sudden onset of epigastric pain, nausea, and vomiting, and limited mucosal histologic studies demonstrate a marked infiltrate of neutrophils with edema and hyperemia. If not treated, this picture will evolve into one of chronic gastritis. Hypochlorhydria lasting for up to 1 year may follow acute *H. pylori* infection.

Bacterial infection of the stomach or phlegmonous gastritis is a rare, potentially life-threatening disorder characterized by marked and diffuse acute inflammatory infiltrates of the entire gastric wall, at times accompanied by necrosis. Elderly individuals, alcoholics, and AIDS patients may be affected. Potential iatrogenic causes include polypectomy and mucosal injection with India ink. Organisms associated with this entity include streptococci, staphylococci, *Escherichia coli*, *Proteus*, and *Haemophilus* species. Failure of supportive measures and antibiotics may result in gastrectomy.

Other types of infectious gastritis may occur in immunocompromised individuals such as AIDS patients. Examples include herpetic (herpes simplex) or CMV gastritis. The histologic finding of intranuclear inclusions would be observed in the latter.

Chronic Gastritis Chronic gastritis is identified histologically by an inflammatory cell infiltrate consisting primarily of lymphocytes and plasma cells, with very scant neutrophil involvement. Distribution of the inflammation may be patchy, initially involving superficial and glandular portions of the gastric mucosa. This picture may progress to more severe glandular destruction, with atrophy and metaplasia. Chronic gastritis has been classified according to histologic characteristics. These include superficial atrophic changes and gastric atrophy. The association of atrophic gastritis with the development of gastric cancer has led to the development of endoscopic and serologic markers of severity. Some of these include gross inspection and classification of mucosal abnormalities during standard endoscopy, magnification endoscopy, endoscopy with narrow band imaging and/or autofluorescence imaging, and measurement of several serum biomarkers including pepsinogen I and II levels, gastrin-17, and anti-*H. pylori* serologies. The clinical utility of these tools is currently being explored.

The early phase of chronic gastritis is *superficial gastritis*. The inflammatory changes are limited to the lamina propria of the surface mucosa, with edema and cellular infiltrates separating intact gastric glands. The next stage is *atrophic gastritis*. The inflammatory infiltrate extends deeper into the mucosa, with progressive distortion and destruction of the glands. The final stage of chronic gastritis is *gastric atrophy*. Glandular structures are lost, and there is a paucity of inflammatory infiltrates. Endoscopically, the mucosa may be substantially thin, permitting clear visualization of the underlying blood vessels.

Gastric glands may undergo morphologic transformation in chronic gastritis. Intestinal metaplasia denotes the conversion of gastric glands to a small intestinal phenotype with small-bowel mucosal glands containing goblet cells. The metaplastic changes may vary in distribution from patchy to fairly extensive gastric involvement. Intestinal metaplasia is an important predisposing factor for gastric cancer (Chap. 80).

Chronic gastritis is also classified according to the predominant site of involvement. Type A refers to the body-predominant form (autoimmune), and type B is the antral-predominant form (*H. pylori*-related). This classification is artificial in view of the difficulty in distinguishing between these two entities. The term *AB gastritis* has been used to refer to a mixed antral/body picture.

TYPE A GASTRITIS The less common of the two forms involves primarily the fundus and body, with antral sparing. Traditionally, this form of gastritis has been associated with pernicious anemia (Chap. 95) in the presence of circulating antibodies against parietal cells and IF; thus, it is also called *autoimmune gastritis*. *H. pylori* infection can lead to a similar distribution of gastritis. The characteristics of an autoimmune picture are not always present.

Antibodies to parietal cells have been detected in >90% of patients with pernicious anemia and in up to 50% of patients with type A gastritis. The parietal cell antibody is directed against H⁺K⁺-ATPase. T cells are also implicated in the injury pattern of this form of gastritis. A subset of patients infected with *H. pylori* develop antibodies against

H^+ , K^+ -ATPase, potentially leading to the atrophic gastritis pattern seen in some patients infected with this organism. The mechanism is thought to involve molecular mimicry between *H. pylori* LPS and H^+ , K^+ -ATPase.

Parietal cell antibodies and atrophic gastritis are observed in family members of patients with pernicious anemia. These antibodies are observed in up to 20% of individuals aged >60 and in ~20% of patients with vitiligo and Addison's disease. About one-half of patients with pernicious anemia have antibodies to thyroid antigens, and ~30% of patients with thyroid disease have circulating anti-parietal cell antibodies. Anti-IF antibodies are more specific than parietal cell antibodies for type A gastritis, being present in ~40% of patients with pernicious anemia. Another parameter consistent with this form of gastritis being autoimmune in origin is the higher incidence of specific familial histocompatibility haplotypes such as HLA-B8 and HLA-DR3. Low pepsinogen levels have also been observed; thus, this marker has been used as an additional diagnostic tool in autoimmune gastritis.

The parietal cell-containing gastric gland is preferentially targeted in this form of gastritis, and achlorhydria results. Parietal cells are the source of IF, the lack of which will lead to vitamin B_{12} deficiency and its sequelae (megaloblastic anemia, neurologic dysfunction).

Gastric acid plays an important role in feedback inhibition of gastrin release from G cells. Achlorhydria, coupled with relative sparing of the antral mucosa (site of G cells), leads to hypergastrinemia. Gastrin levels can be markedly elevated (>500 pg/mL) in patients with pernicious anemia. ECL cell hyperplasia with frank development of gastric carcinoid tumors may result from gastrin trophic effects. Hypergastrinemia and achlorhydria may also be seen in nonpernicious anemia-associated type A gastritis.

TYPE B GASTRITIS Type B, or antral-predominant, gastritis is the more common form of chronic gastritis. *H. pylori* infection is the cause of this entity. Although described as "antral-predominant," this is likely a misnomer in view of studies documenting the progression of the inflammatory process toward the body and fundus of infected individuals. The conversion to a pangastritis is time dependent and estimated to require 15–20 years. This form of gastritis increases with age, being present in up to 100% of persons aged >70. Histology improves after *H. pylori* eradication. The number of *H. pylori* organisms decreases dramatically with progression to gastric atrophy, and the degree of inflammation correlates with the level of these organisms. Early on, with antral-predominant findings, the quantity of *H. pylori* is highest and a dense chronic inflammatory infiltrate of the lamina propria is noted, accompanied by epithelial cell infiltration with polymorphonuclear leukocytes (Fig. 324-16).

Multifocal atrophic gastritis, gastric atrophy with subsequent metaplasia, has been observed in chronic *H. pylori*-induced gastritis. This may ultimately lead to development of gastric adenocarcinoma

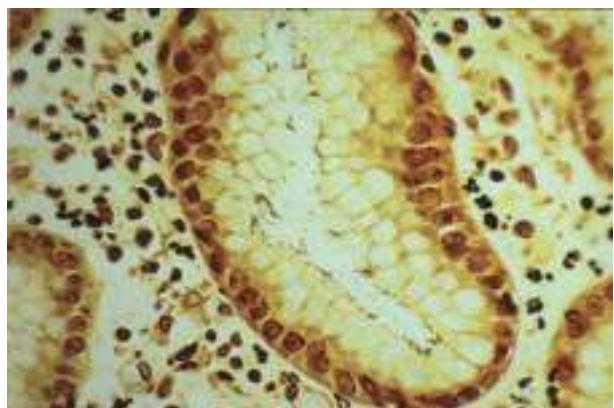


FIGURE 324-16 Chronic gastritis and *H. pylori* organisms. Steiner silver stain of superficial gastric mucosa showing abundant darkly stained microorganisms layered over the apical portion of the surface epithelium. Note that there is no tissue invasion.

(Fig. 324-8; Chap. 80). *H. pylori* infection is now considered an independent risk factor for gastric cancer. Worldwide epidemiologic studies have documented a higher incidence of *H. pylori* infection in patients with adenocarcinoma of the stomach as compared to control subjects. Seropositivity for *H. pylori* is associated with a three- to sixfold increased risk of gastric cancer. This risk may be as high as ninefold after adjusting for the inaccuracy of serologic testing in the elderly. The mechanism by which *H. pylori* infection leads to cancer is unknown, but it appears to be related to the chronic inflammation induced by the organism. Eradication of *H. pylori* as a general preventative measure for gastric cancer is being evaluated but is not yet recommended.

Infection with *H. pylori* is also associated with development of a low-grade B-cell lymphoma, gastric MALT lymphoma (Chap. 108). The chronic T-cell stimulation caused by the infection leads to production of cytokines that promote the B-cell tumor. The tumor should be initially staged with a CT scan of the abdomen and EUS. Tumor growth remains dependent on the presence of *H. pylori*, and its eradication is often associated with complete regression of the tumor. The tumor may take more than a year to regress after treating the infection. Such patients should be followed by EUS every 2–3 months. If the tumor is stable or decreasing in size, no other therapy is necessary. If the tumor grows, it may become a high-grade B-cell lymphoma. When the tumor becomes a high-grade aggressive lymphoma histologically, it loses responsiveness to *H. pylori* eradication.

TREATMENT

Chronic Gastritis

Treatment in chronic gastritis is aimed at the sequelae and not the underlying inflammation. Patients with pernicious anemia will require parenteral vitamin B_{12} supplementation on a long-term basis. Eradication of *H. pylori* is often recommended even if PUD or a low-grade MALT lymphoma is not present. Expert opinion suggests that patients with atrophic gastritis complicated by intestinal metaplasia without dysplasia should undergo surveillance endoscopy every 3 years.

Miscellaneous Forms of Gastritis *Lymphocytic gastritis* is characterized histologically by intense infiltration of the surface epithelium with lymphocytes. The infiltrative process is primarily in the body of the stomach and consists of mature T cells and plasmacytoid dendritic cells. The etiology of this form of chronic gastritis is unknown. It has been described in patients with celiac sprue, but whether there is a common factor associating these two entities is unknown. No specific symptoms suggest lymphocytic gastritis. A subgroup of patients has thickened folds noted on endoscopy. These folds are often capped by small nodules that contain a central depression or erosion; this form of the disease is called *varioliform gastritis*. *H. pylori* probably plays no significant role in lymphocytic gastritis. Therapy with glucocorticoids or sodium cromoglycate has obtained unclear results.

Marked eosinophilic infiltration involving any layer of the stomach (mucosa, muscularis propria, and serosa) is characteristic of *eosinophilic gastritis*. Affected individuals will often have circulating eosinophilia with clinical manifestation of systemic allergy. Involvement may range from isolated gastric disease to diffuse eosinophilic gastroenteritis. Antral involvement predominates, with prominent edematous folds being observed on endoscopy. These prominent antral folds can lead to outlet obstruction. Patients can present with epigastric discomfort, nausea, and vomiting. Treatment with glucocorticoids has been successful.

Several systemic disorders may be associated with *granulomatous gastritis*. Gastric involvement has been observed in Crohn's disease. Involvement may range from granulomatous infiltrates noted only on gastric biopsies to frank ulceration and stricture formation. Gastric Crohn's disease usually occurs in the presence of small-intestinal disease. Several rare infectious processes can lead to granulomatous gastritis, including histoplasmosis, candidiasis, syphilis, and tuberculosis.

Other unusual causes of this form of gastritis include sarcoidosis, idiopathic granulomatous gastritis, and eosinophilic granulomas involving the stomach. Establishing the specific etiologic agent in this form of gastritis can be difficult, at times requiring repeat endoscopy with biopsy and cytology. Occasionally, a surgically obtained full-thickness biopsy of the stomach may be required to exclude malignancy.

Russell body gastritis (RBG) is a mucosal lesion of unknown etiology that has a pseudotumoral endoscopic appearance. Histologically, it is defined by the presence of numerous plasma cells containing Russell bodies (RBs) that express kappa and lambda light chains. Only 10 cases have been reported, and 7 of these have been associated with *H. pylori* infection. The lesion can be confused with a neoplastic process, but it is benign in nature, and the natural history of the lesion is not known. There have been cases of resolution of the lesion when *H. pylori* was eradicated.

Immune checkpoint inhibitor-induced enterocolitis and gastritis are recognized sequelae of these oncologic therapies. The gastritis typically occurs later in the course of therapy. The diagnosis is made by the histologic findings on gastric mucosal biopsies obtained endoscopically. This is an important diagnosis to make since therapy with glucocorticoids and potentially IL-6 receptor blockers will be required. Moreover, this side effect will have an effect on the oncologic therapy prescribed.

MÉNÉTRIER'S DISEASE

Ménétrier's disease (MD) is a very rare gastropathy characterized by large, tortuous mucosal folds. MD has an average age of onset of 40–60 years with a male predominance. The differential diagnosis of large gastric folds includes ZES, malignancy (lymphoma, infiltrating carcinoma), infectious etiologies (CMV, histoplasmosis, syphilis, tuberculosis), gastritis polyposa profunda, and infiltrative disorders such as sarcoidosis. MD is most commonly confused with large or multiple gastric polyps (prolonged PPI use) or familial polyposis syndromes. The mucosal folds in MD are often most prominent in the body and fundus, sparing the antrum. Histologically, massive foveolar hyperplasia (hyperplasia of surface and glandular mucous cells) and a marked reduction in oxytic glands and parietal cells and chief cells are noted. This hyperplasia produces the prominent folds observed. The pits of the gastric glands elongate and may become extremely dilated and tortuous. Although the lamina propria may contain a mild chronic inflammatory infiltrate including eosinophils and plasma cells, MD is not considered a form of gastritis. The etiology of this unusual clinical picture in children is often CMV, but the etiology in adults is unknown. Overexpression of the growth factor TGF- α has been demonstrated in patients with MD. The overexpression of TGF- α in turn results in overstimulation of the epidermal growth factor receptor (EGFR) pathway and increased proliferation of mucus cells, resulting in the observed foveolar hyperplasia.

The clinical presentation in adults is usually insidious and progressive. Epigastric pain, nausea, vomiting, anorexia, peripheral edema, and weight loss are signs and symptoms in patients with MD. Occult GI bleeding may occur, but overt bleeding is unusual and, when present, is due to superficial mucosal erosions. In fact, bleeding is more often seen in one of the common mimics of MD, gastric polyposis. Twenty to 100% of patients (depending on time of presentation) develop a protein-losing gastropathy due to hypersecretion of gastric mucus accompanied by hypoalbuminemia and edema. Gastric acid secretion is usually reduced or absent because of the decreased parietal cells. Large gastric folds are readily detectable by either radiographic (barium meal) or endoscopic methods. Endoscopy with deep mucosal biopsy, preferably full thickness with a snare technique, is required to establish the diagnosis and exclude other entities that may present similarly. A nondiagnostic biopsy may lead to a surgically obtained full-thickness biopsy to exclude malignancy. Although MD is considered premalignant by some, the risk of neoplastic progression is not defined. Complete blood count, serum gastrin, serum albumin, CMV and *H. pylori* serology, and pH testing of gastric aspirate during endoscopy should be included as part of the initial evaluation of patients with large gastric folds.

TREATMENT

Ménétrier's Disease

Medical therapy with anticholinergic agents, prostaglandins, PPIs, prednisone, somatostatin analogues (octreotide), and H₂ receptor antagonists yields varying results. Ulcers should be treated with a standard approach. The discovery that MD is associated with overstimulation of the EGFR pathway has led to the successful use of the EGF inhibitory antibody, cetuximab, in these patients. Specifically, four of seven patients who completed a 1-month trial with this agent demonstrated near complete histologic remission and improvement in symptoms. Cetuximab is now considered the first-line treatment for MD, leaving partial or total gastrectomy for severe disease with persistent and substantial protein loss despite therapy with this agent.

FURTHER READING

- B S et al: Non-steroidal anti-inflammatory drugs (NSAIDs) and organ damage: A current perspective. *Biochem Pharmacol* 180:114147, 2020.
- B I et al: Mechanisms of damage to the gastrointestinal tract from nonsteroidal anti-inflammatory drugs. *Gastroenterology* 154:500, 2018.
- B ML et al: Multiple endocrine neoplasia type 1: Latest insights. *Endocr Rev* 42:133, 2021.
- C WD et al: ACG clinical guideline: Treatment of *Helicobacter pylori* infection. *Am J Gastroenterol* 112:212, 2017.
- E AC et al: The physiology of the gastric parietal cell. *Physiol Rev* 100:573, 2019.
- J RT, I T: Gastrinoma; Endotext [internet]. South Dartmouth, MA, 2020. <https://europepmc.org/article/NBK/nbk279075>.
- K RT et al: Diagnosis and treatment of peptic ulcer disease. *Am J Med* 132:447, 2019.
- P G et al: Gastritis: Update on etiological features and histological practice approach. *Pathologica* 112:153, 2020.
- S V et al: Proton pump inhibitors: Use and misuse in the clinical setting. *Expert Rev Clin Pharmacol* 11:1123, 2018.
- Y X, S AJ: Gastric parietal cell physiology and *Helicobacter pylori*-induced disease. *Gastroenterology* 156:2158, 2019.

325

Disorders of Absorption

Deborah C. Rubin



A wide range of diseases affect gastrointestinal (GI) absorptive function and may result in malabsorption syndromes. These disorders affect one or more of the three phases of enteral nutrient processing. Luminal digestion is initiated by lingual and gastric lipase and gastric pepsin, and continues in the small bowel by the actions of pancreatic enzymes and bile salts. Small intestinal mucosal digestion and absorption are mediated by enterocyte brush border enzymes including disaccharidases, enterokinases, and peptidases, which digest nutrients upon contact, and by mixed micelles containing lipids and bile salts. Protein and carbohydrate digestive products are transported into the enterocyte by carriers and transporters, and lipids enter by diffusion mediated by micelles. Once in the enterocyte, nutrients may be reprocessed for post mucosal absorption and entry into lymphatics (long-chain triglycerides as part of chylomicrons) or are transported into the bloodstream. Malabsorptive diseases or syndromes can be classified by their effects on one or more of these three phases of absorption (Table 325-1).

TABLE 325-1 Classification of Malabsorption Syndromes

Inadequate digestion
Postgastrectomy ^a
Deficiency or inactivation of pancreatic lipase
Exocrine pancreatic insufficiency
Chronic pancreatitis
Pancreatic carcinoma
Cystic fibrosis
Pancreatic insufficiency—congenital or acquired
Gastrinoma—acid inactivation of lipase
Drugs—orlistat
Reduced intraduodenal bile-acid concentration/impaired micelle formation
Liver disease
Parenchymal liver disease
Cholestatic liver disease
Bacterial overgrowth in small intestine:
Anatomic stasis
Afferent loop
Stasis/blind
Loop/strictures/fistulae
Functional stasis
Diabetes ^a
Scleroderma ^a
Intestinal pseudo-obstruction
Interrupted enterohepatic circulation of bile salts
Ileal resection
Crohn's disease
Drugs (binding or precipitating bile salts)—neomycin, cholestyramine, calcium carbonate
Impaired mucosal absorption/mucosal loss or defect
Intestinal resection or bypass ^a
Inflammation, infiltration, or infection:
Crohn's disease ^a
Amyloidosis
Scleroderma ^a
Lymphoma ^a
Eosinophilic enteritis
Mastocytosis
Tropical sprue
Celiac disease
Collagenous sprue
Whipple's disease ^a
Radiation enteritis ^a
Folate and vitamin B ₁₂ deficiency
Infections—giardiasis
Graft vs host disease
Genetic disorders
Disaccharidase deficiency
Agammaglobulinemia
Abetalipoproteinemia
Hartnup disease
Cystinuria
Impaired nutrient delivery to and/or from intestine:
Lymphatic obstruction
Lymphoma ^a
Lymphangiectasia
Circulatory disorders
Congestive heart failure
Constrictive pericarditis
Mesenteric artery atherosclerosis
Vasculitis
Endocrine and metabolic disorders
Diabetes ^a
Hypoparathyroidism
Adrenal insufficiency
Hyperthyroidism
Carcinoid syndrome

^aMalabsorption caused by more than one mechanism.

Disorders of absorption also have diverse clinical presentations. For example, the deficiency of a single brush border membrane protein such as lactase causes symptoms of diarrhea by affecting the absorption of one nutrient, lactose. Celiac sprue may be localized to the duodenum

and present with isolated iron deficiency, or may cause diffuse intestinal mucosal disease, affecting the absorption of multiple nutrients and causing a constellation of symptoms and clinical presentations.

Definition of Diarrhea Diarrhea is the most common symptom associated with disorders of absorption. For most patients, diarrhea as a symptom is defined as an increase in stool number or frequency, or a change in consistency. Because normal bowel patterns may vary from as many as two to four bowel movements per day to one stool per week, it is critical to use an objective measure of diarrhea to help direct evaluation. In health, stool volume or weight is <200 mL or <200 g respectively in 24 h. Collection of stool for weight/volume determination is one of the most useful tools for an evaluation of diarrhea. In particular, a 72-h collection for weight/volume and fecal fat determination is the gold standard for documenting the presence of steatorrhea, or fatty stool. Steatorrhea, defined as increased stool fat excretion to >7% of dietary fat, is a common manifestation of malabsorption. Steatorrhea often results in large, bulky, and malodorous stools. Malabsorption of single nutrients like lactose may result in an osmotic diarrhea, in which the osmotically active unabsorbed nutrient causes fluid to be drawn into the GI tract lumen. Malabsorptive diarrhea frequently is precipitated by eating and resolves or significantly decreases at night, with fasting, and thus can frequently be distinguished from secretory diarrheas, for example from infectious causes such as bacterial enterotoxigenic *Escherichia coli*. In this circumstance, intestinal fluid and electrolyte secretion is stimulated by enterotoxin and will continue even during fasting.

OVERVIEW: NUTRIENT DIGESTION AND ABSORPTION

Luminal digestive processes begin in the mouth and proceed throughout the GI tract, mediated by salivary amylase, lingual and gastric lipases, gastric acid, pancreatic enzymes, and bile salts. As nutrients are digested in the lumen of the proximal GI tract, they are further processed by enterocyte brush border enzymes including disaccharidases such as lactase and sucrase-isomaltase, which produce monosaccharides, and peptidases, which hydrolyze polypeptides into tripeptides and dipeptides and amino acids. Lipids in mixed micelles are then absorbed into enterocytes.

The surface area of the small bowel, which is normally 6–12 ft long, is further enhanced by circular folds, villi, and microvilli. Following uptake into enterocytes, nutrients are further processed and transported into the lymphatics or into the portal circulation for use by other cells throughout the body. The intestine is also presented with 7–9 L of fluid daily, a volume comprising dietary fluid intake (1–2 L/day) and salivary, gastric, pancreatic, biliary, and intestinal fluid (6–7 L/day). In health, almost all of this fluid is reabsorbed by the small bowel and colon, resulting in a normal stool volume of <200 mL or stool weight of <200 g.

SPECIFIC NUTRIENTS

Lipids Lipid absorption is a complex process that requires hydrolysis by pancreatic enzymes and bile salts for physiochemical dispersion of fats, followed by absorption of processed lipid nutrients dispersed in bile salt-mixed micelles across the intestinal epithelium. Bile acids are synthesized in the liver, secreted into the intestinal lumen, and constantly recirculated by absorption in the ileum. The ileum expresses fibroblast growth factor 19 (FGF19), which is a physiologic bile acid sensor. FGF19 is secreted from the ileum into the bloodstream in response to bile acid flux and negatively regulates hepatic bile acid synthesis by affecting the transcription of hepatic CYPTA1.

Thus assimilation of dietary lipid requires three integrated processes: an intraluminal or digestive phase, a mucosal or absorptive phase, and a delivery or postabsorptive phase (Table 325-2).

Gastric lipases begin the lipolytic process. Following entry into the small bowel, long-chain triglycerides, with carbon lengths >12 and that are the major component of dietary lipid, are hydrolyzed by pancreatic lipases into fatty acids and monoglyceride during a process

TABLE 325-2 Defects in Lipid Digestion and Absorption in Steatorrhea

PHASE/PROCESS	PATHOPHYSIOLOGIC DEFECT	DISEASE EXAMPLE
Digestive		
Lipolysis formation	Decreased lipase secretion	Chronic pancreatitis
Micelle formation	Decreased intraduodenal bile acids	
Absorptive		
Mucosal uptake and re-esterification	Mucosal dysfunction	Celiac disease
Postabsorptive		
Chylomicron formation	Absent β -lipoproteins	Abetalipoproteinemia
Delivery from intestine	Abnormal lymphatics	Intestinal lymphangiectasia

called *lipolysis* (Fig. 325-1). Long-chain free fatty acids are dispersed by bile salts into mixed micelles, which contact the brush border and permit fatty acid absorption into enterocytes across this specialized apical membrane. The other two types of fatty acids that compose fats, medium-chain and short-chain fatty acids, are soluble in the unstirred water layer. Medium-chain triglycerides with carbon chain lengths of 8–12 are found in coconut oil. Long-chain fatty acids are re-esterified to triglycerides in enterocytes, packaged into chylomicrons that contain apolipoproteins on the surface, which are subsequently secreted into the extracellular space, and because of their size, are excluded from capillaries and enter the lymphatics. Medium-chain triglycerides do not require micelle formation or pancreatic lipolysis as they are directly absorbed intact from the small bowel into the bloodstream, and short-chain fatty acids (carbon length <8) are produced by and absorbed in the colon.

Carbohydrates Dietary carbohydrate consists of starch, sucrose, lactose, maltose, and monosaccharides such as glucose and fructose. Starch is digested by salivary α -amylase in the mouth, followed by pancreatic amylase. The main products include maltotriose, maltose, and α -dextrins. These are further digested on the brush border membrane by disaccharidases such as glucoamylase and sucrase-isomaltase. Dietary lactose is digested by brush border lactase, sucrose by sucrase, and trehalose by trehalase. The final digested products are glucose, fructose, and galactose, which are transported into the enterocyte by transporters such as SLCA5 (formerly SGLT-1), which transports glucose or galactose in a sodium-dependent manner, and GLUT-5, which transports fructose by facilitated diffusion. Glucose, galactose, and fructose exit the cell via GLUT-2.

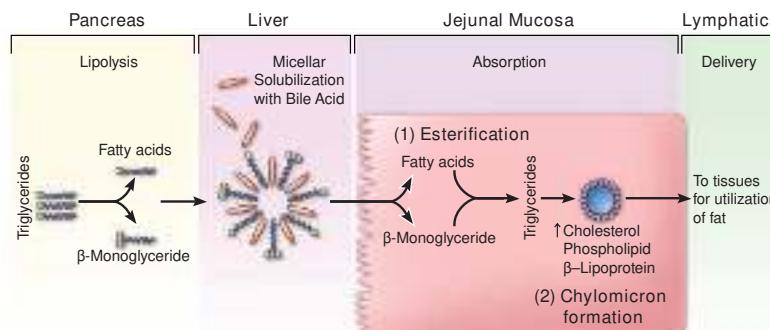


FIGURE 325-1 Schematic representation of lipid digestion and absorption. Dietary lipid is in the form of long-chain triglycerides. The overall process can be divided into (1) a digestive phase that includes both lipolysis and micelle formation requiring pancreatic lipase and conjugated bile acids, respectively, in the duodenum; (2) an absorptive phase for mucosal uptake and re-esterification; and (3) a postabsorptive phase that includes chylomicron formation and exit from the intestinal epithelial cell via lymphatics. (Courtesy of John M. Dietschy, MD; with permission.)

Proteins Dietary protein digestion begins in the stomach by pepsin. Pancreatic proteases including endopeptidases, exopeptidases, and trypsin are activated in the small-bowel lumen. Trypsinogen is activated by brush border enterokinase to generate active trypsin. Trypsin in turn activates chymotrypsinogen to chymotrypsin, proelastase to elastase, and procarboxypeptidases to carboxypeptidases A and B. These enzymes digest protein into di peptides, tripeptides, larger polypeptides, or free amino acids. At the brush border, peptidases digest larger peptides into dipeptides and tripeptides or free amino acids, which enter the enterocyte via specialized carriers. Most dipeptides and tripeptides are further metabolized intracellularly by cytoplasmic peptidase into amino acids, which directly enter the bloodstream via carriers in the basolateral membrane. Small amounts of dipeptides and tripeptides may also enter the bloodstream.

LUMINAL PHASE OF DIGESTION

The luminal phase of digestion begins in the mouth, starting with mastication and lipase secretion by the tongue and salivary glands. The stomach continues the luminal digestive process, via gastric acid, gastric lipase, and pepsin secretion as well as mechanical trituration of contents. In the small-bowel lumen, pancreatic enzymes (amylase, lipases, carboxypeptidase, trypsin, and other endopeptidases) contribute to carbohydrate, lipid, and protein digestion, respectively. Bile salts produced by the liver are secreted into the intestinal lumen (and reabsorbed in the ileum via the enterohepatic circulation) and are required for efficient lipid absorption.

Disorders That Affect the Luminal Phase of Digestion The luminal phase may be disrupted by disorders of gastric and intestinal motility including the sequelae of gastric surgery, systemic diseases such as scleroderma, or endocrine disorders such as diabetes mellitus, pancreatic diseases leading to pancreatic insufficiency with reduced pancreatic enzyme secretion, or luminal bile salt deficiency caused by hepatobiliary disease, ileal disease, or small-bowel bacterial overgrowth.

Gastric Resection Surgical procedures that remove or bypass part of the stomach and duodenal bulb such as Roux-en-Y gastric bypass for weight loss, or resection of the gastric antrum and duodenal bulb with creation of a Billroth II anastomosis for treatment of peptic ulcer disease, result in rapid gastric emptying into the jejunum, which leads to diarrhea and weight loss due to inadequate mixing of luminal nutrients with bile and pancreatic secretions.

Disordered Intestinal Motility Hyperthyroidism may cause diarrhea and malabsorption due to increased intestinal motility with rapid transit, also resulting in inadequate nutrient mixing with pancreaticobiliary secretions. Long-standing diabetes mellitus may result

in damage to the enteric nervous system resulting in increased motility and diarrhea, or reduced motility and constipation. Disorders that affect the intestinal smooth muscle such as connective tissue disorders including scleroderma may have profound effects on GI motility.

Pancreatic Disorders Chronic pancreatitis (see Chap. 348) may result in a marked reduction in pancreatic enzyme secretion and pancreatic insufficiency, with subsequent fat, protein, and carbohydrate malabsorption. Patients with chronic pancreatitis present with steatorrhea, or fatty stools, which are often voluminous, bulky, and malodorous. Patients with steatorrhea also develop deficiency of fat-soluble vitamins including vitamins A, E, and most commonly, vitamins D and K, which depend on the same lipid absorption mechanisms, and thus are malabsorbed along with dietary fat. Weight loss is common. For a discussion of causes of acute and chronic pancreatitis, please see Chap. 348.

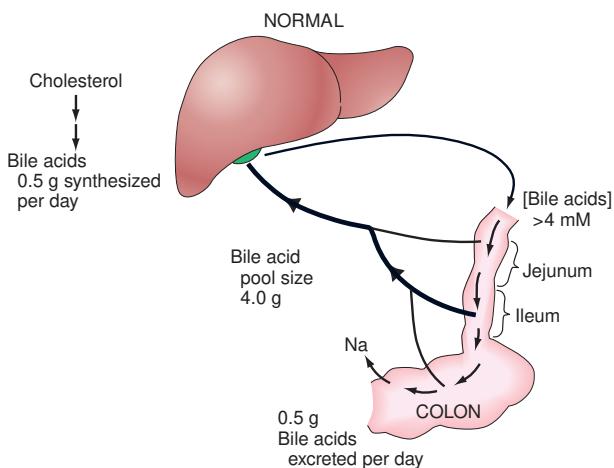


FIGURE 325-2 Schematic representation of the enterohepatic circulation of bile acids. Bile-acid synthesis is cholesterol catabolism and occurs in the liver. Bile acids are secreted in bile and are stored in the gallbladder between meals and at night. Food in the duodenum induces the release of cholecystokinin, a potent stimulus for gallbladder contraction resulting in bile-acid entry into the duodenum. Bile acids are primarily absorbed via an Na^+ -dependent transport process that is located only in the ileum. A relatively small quantity of bile acids ($\sim 500 \text{ mg}$) is not absorbed in a 24-h period and is lost in stool. Fecal bile-acid losses are matched by bile-acid synthesis. The bile-acid pool (the total amount of bile acids in the body) is $\sim 4 \text{ g}$ and is circulated twice during each meal or six to eight times in a 24-h period.

Disorders That Result in Luminal Bile Salt Deficiency Bile acid synthesis and the enterohepatic circulation (Fig. 325-2): Bile acids are synthesized from cholesterol in the liver. The two primary bile acids are cholic acid and chenodeoxycholic acid. These are conjugated in the liver to taurine and glycine and are secreted into bile ducts, stored in the gallbladder, and then delivered to the intestinal lumen. Conjugation prevents bile acids from passive diffusion in the small-bowel lumen, retaining bile acid concentrations required for lipid absorption. Bile acids emulsify fats and fat-soluble vitamins to facilitate their absorption. Bile acids are efficiently reabsorbed in the ileum into the portal circulation and are extracted by the liver in a process called *enterohepatic circulation* (Fig. 325-2). Small amounts are deconjugated in the ileum by bacteria, or pass into the colon and are deconjugated and metabolized by colonic bacteria to become secondary bile acids. The two major secondary bile acids are lithocholic acid and deoxycholic acid.

Processes that affect any of the above pathways may result in luminal bile salt deficiency and malabsorption. Thus, hepatobiliary diseases, intestinal ileal resection, extensive disease such as Crohn's disease, and small-bowel bacterial overgrowth may result in luminal bile salt deficiency and malabsorption (Table 325-3).

Hepatobiliary Disease Hepatic disorders that result in decreased bile acid synthesis due to hepatocyte dysfunction or reduced secretion of bile into the gut lumen caused by diseases of the bile ducts such as primary sclerosing cholangitis or primary biliary cirrhosis may result in luminal bile salt deficiency and fat malabsorption. These are discussed in Chap. 346.

TABLE 325-4 Comparison of Bile Acid and Fatty Acid Diarrhea

	BILE ACID DIARRHEA	FATTY ACID DIARRHEA
Extent of ileal disease	Limited	Extensive
Ileal bile-acid absorption	Reduced	Reduced
Fecal bile-acid excretion	Increased	Increased
Fecal bile-acid loss compensated by hepatic synthesis	Yes	No
Bile-acid pool size	Normal	Reduced
Intraduodenal (bile acid)	Normal	Reduced
Steatorrhea	None or mild	$>20 \text{ g}$
Response to cholestyramine	Yes	No
Response to low-fat diet	No	Yes

Ileal Resection or Ileal Disease Diseases that involve the ileal mucosa or that result in ileal resection may lead to reduced recycling of bile acids by the enterohepatic circulation and increased entry into and concentration of bile acids in the colon, which produces a secretory diarrhea, or malabsorption due to inadequate bile acid concentrations in the small-bowel lumen. In general, resection or disease involving $<100 \text{ cm}$ of ileum results in bile acid spillage into the colon; resections of $>100 \text{ cm}$ result in loss of bile acids that exceed liver synthetic capacity, and malabsorption becomes the dominant pathophysiologic mechanism for diarrhea, due to bile acid deficiency (Table 325-4). The most common disorder of the GI tract that targets the ileum is Crohn's disease (Chap. 326), which is a chronic inflammatory disorder that may involve the entire GI tract, but most commonly the ileum and colon. If severe or refractory to treatment, Crohn's disease may lead to chronic inflammation, marked epithelial dysfunction, and structuring and fibrosis, and surgical resection may be required to treat small-bowel obstruction or refractory disease.

Primary Bile Acid Diarrhea A subset of patients with functional diarrhea or irritable bowel syndrome with diarrhea have been recently shown to have bile acid malabsorption. Although the mechanisms are still being elucidated, reduced FGF19 secretion by ileal enterocytes has been observed. FGF19 regulates serum 7alpha-hydroxy-4-cholesteno-3-one (C4) levels; reductions in circulating FGF19 lead to increased hepatic bile acid synthesis via increased C4 expression. Chronic diarrhea results from increased bile acid spillage into the colon, which induces a secretory diarrhea.

Treatment Bile acid sequestrants are effective in reducing diarrhea by binding bile acids to prevent spillage into the colon. Hepatic synthesis of bile acids is sufficient to maintain intraluminal concentrations that are adequate for fat absorption.

Small-Bowel Bacterial Overgrowth The intestine contains a rich microbiome. Bacterial titers increase along the horizontal axis of the gut from duodenum to ileum. However, intestinal disorders affecting motility or causing stasis of bowel contents may lead to small-bowel bacterial overgrowth. These include scleroderma bowel, chronic intestinal pseudo-obstruction, the creation of blind surgical loops such as Billroth II anastomosis, small-bowel strictures, or fibrosis from inflammatory disorders such as Crohn's disease, and diffuse diverticulosis (Fig. 325-3). Surgical resection of the ileocecal valve increases ileal bacterial counts from the colon. Bacterial overgrowth causes deconjugation of bile acids, which facilitates their absorption in the proximal bowel and results in luminal bile acid deficiency, which in turn causes malabsorptive diarrhea with steatorrhea. Bacterial overgrowth may also damage the brush border and result in carbohydrate maldigestion and short-chain fatty acid production in the colon, with diarrhea and gas. These patients are also at risk for B_{12} deficiency due to bacterial metabolism of B_{12} , resulting in macrocytic anemia and peripheral neuropathy. In contrast, elevated serum folate levels may also be observed, derived from bacterial synthesis of folate.

Small-bowel bacterial overgrowth has also been observed in patients with diarrhea-predominant irritable bowel syndrome. The underlying

TABLE 325-3 Defects in Enterohepatic Circulation of Bile Acids

PROCESS	PATHOPHYSIOLOGIC DEFECT	DISEASE EXAMPLE
Synthesis	Decreased hepatic function	Cirrhosis
Biliary secretion	Altered canalicular function	Primary biliary cirrhosis
Maintenance of conjugated bile acids	Bacterial overgrowth	Jejunal diverticulosis
Reabsorption	Abnormal ileal function	Crohn's disease

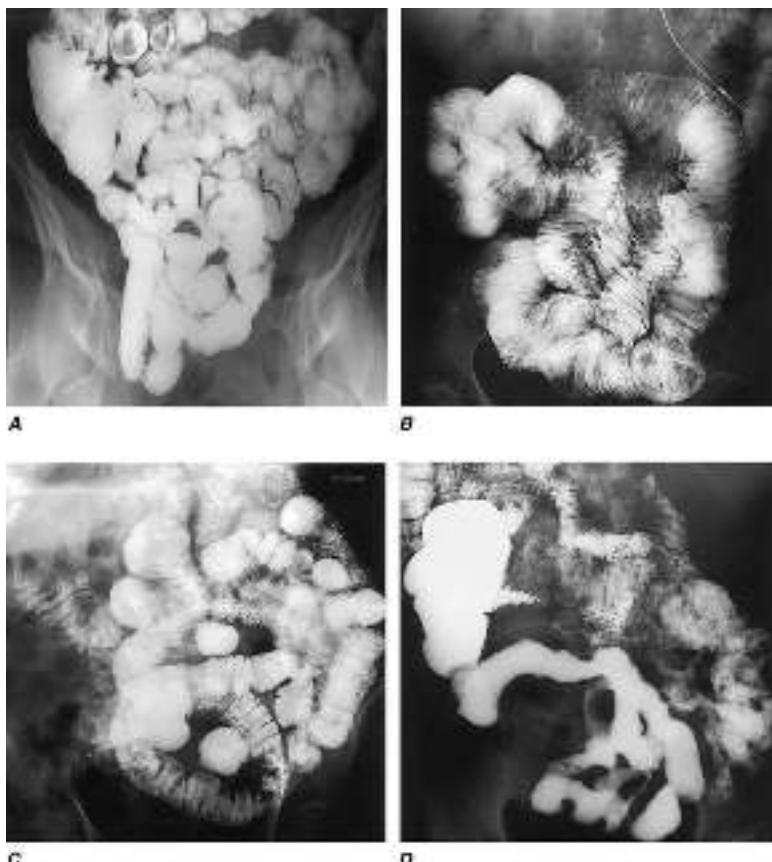


FIGURE 325-3 Barium contrast small-intestinal radiologic examinations. *A*. Normal individual. *B*. Celiac disease. *C*. Jejunal diverticulosis. *D*. Crohn's disease. (Courtesy of Morton Burrell, MD, Yale University; with permission.)

mechanisms are unclear, but treatment of bacterial overgrowth leads to resolution of symptoms in a subset of irritable bowel syndrome patients.

Diagnosis Duodenal aspirate for bacterial titers is the gold standard but is not generally available to most practitioners. Breath hydrogen testing with administration of lactulose, a nondigestible disaccharide, is widely available but must be interpreted carefully to avoid false-positive results. Many clinicians choose to treat empirically with antibiotics (see Treatment) and observe for resolution of symptoms.

Treatment When possible, surgical correction of blind loops, endoscopic or surgical treatment of strictures, and removal of large diverticula can be pursued for definitive therapy, in addition to treatment of underlying disorders such as Crohn's disease to avoid recurrent stricture formation or fibrosis. Other disorders such as scleroderma or other diffuse motility disorders may not be easily treated. In these circumstances, treatment with the nonabsorbable antibiotic, rifaximin, or with other antibiotics such as metronidazole, doxycycline, amoxicillin-clavulanic acid, or cephalosporins for several weeks is often pursued. Patients may require retreatment or even chronic therapy with rotating antibiotics depending on the severity of symptoms.

MUCOSAL PHASE OF DIGESTION AND ABSORPTION

The intestinal epithelium (also known as the mucosa) plays a critical role in continued digestion of nutrients and absorption from the intestinal lumen into the bloodstream and lymphatics.

The small-bowel epithelial or mucosal digestive and absorptive phase is mediated by enterocytic brush border enzymes, including peptidases and hydrolases. Brush border enterokinase is required for the conversion of pancreatic trypsinogen to trypsin, which further

activates trypsinogen and other pancreatic protease proenzymes. The brush border membrane of the small-bowel epithelium expresses a wide variety of disaccharidases, peptidases, and other hydrolases that continue the digestive process for carbohydrates and proteins, with enzymatic digestion of disaccharides to monosaccharides and dipeptidases to amino acids, which are then absorbed by specific transporters. Long-chain fatty acids are re-esterified to triglycerides in enterocytes, packaged into chylomicrons with apolipoproteins on the surface, which are subsequently secreted into the extracellular space, and because of their size, are excluded from capillaries and enter the lymphatics.

INTESTINAL MUCOSAL DISORDERS

■ DISORDERS OF ENTEROCYTE CARBOHYDRATE TRANSPORTERS AND ENZYME DEFICIENCIES

Lactose Intolerance Due to Lactase Deficiency This is the most common brush border disaccharidase deficiency and is a frequent cause of diarrhea, abdominal pain, gassiness, and bloating. Lactose is present in many dairy products but is also a "hidden" component of a vast number of processed foods.

Lactose malabsorption can result from lactase deficiency, which is regulated by primary genetic mechanisms (adult-type hypolactasia) or secondary due to damage to the epithelial (mucosal) lining of the gut, from infections (viral, bacterial, or parasitic) or from intestinal mucosal diseases. Congenital lactase deficiency is very rare and is an autosomal recessive disorder. Hypolactasia in adulthood is very common throughout the world and is considered to be the genetic wild-type; lactase persistence results from a C to T mutation (*LACTASE* *LCT-13910CT* and *LCT-13910TT*) and adults with hypolactasia have absence of this "persistence" allele. Lactose is metabolized by lactase

into glucose and galactose, which are both absorbed by transporters at the enterocyte surface. Patients who are lactase deficient have elevated luminal lactose levels upon ingestion of lactose. The mechanism for diarrhea in lactase deficiency is complex. Undigested lactose acts as an osmotic substance to draw fluid into the small-bowel lumen. In addition, when unabsorbed lactose enters the colon, luminal bacteria ferment lactose producing intestinal gas (hydrogen, carbon dioxide, and methane), bloating, and abdominal pain. Luminal lactose is metabolized by bacteria into short-chain fatty acids that can be absorbed by the colon, but watery diarrhea may occur when a large lactose load exceeds the colon's absorptive capacity.

Diagnosis When lactose intolerance is suspected, a common initial approach is to institute a lactose-exclusion diet and assess for resolution of symptoms. This is a rapid and generally effective diagnostic and therapeutic method. Patients are provided with a list of lactose-containing foods and lactose-free alternatives. Patients are also counseled on alternative calcium sources, because dairy-containing foods are a major source of dietary calcium, which is important for osteoporosis prevention.

Should the results of dietary exclusion be ambiguous, a lactose-tolerance test or breath hydrogen test may prove useful. For the lactose-tolerance test, patients ingest a standardized liquid lactose solution (usually 50 g of lactose) followed by timed measurements of serum glucose for 90 min. If lactose digestion is normal, glucose levels should rise by >20 mg/L. Serum glucose rise <20 mg/L plus the presence of symptoms of lactose intolerance (abdominal discomfort, gassiness, and diarrhea) is considered a positive test. A breath hydrogen test is performed by measuring breath hydrogen levels following ingestion of a standardized lactose load. Breath hydrogen levels should not exceed >20 ppm above the fasting baseline. Generally the peak occurs between 2–4 h. Both methods may be inaccurate if the patient has abnormal gastric emptying or abnormal intestinal transit. Breath hydrogen measurements may be abnormal in the setting of bacterial overgrowth, which may cause very similar symptoms.

Treatment Patients may elect to completely eliminate lactose from their diets. It is very important to consider calcium and vitamin D supplementation because elimination of milk and soft cheeses removes important dietary sources. They also may need to consult a dietitian for guidance about hidden lactose in prepared or other foods. An alternative is to consider using lactase supplementation, which is available over the counter, but which may need to be titrated to avoid symptoms.

Glucose Galactose Malabsorption This rare congenital disorder is an autosomal recessive disease in which mutations occur in the *SLC5A1* gene (also known as *SGLT1*). *SLC5A1* is a brush border protein and member of the sodium-dependent glucose transporter family; mutations in this gene result in malabsorption of glucose and galactose. Gene sequencing has shown that most patients have loss of function single-nucleotide variations. *SLC5A1* actively transports glucose or galactose coupled to sodium cotransport; patients who are homozygous for these loss-of-function variants have severe congenital diarrhea and death if unrecognized. Treatment focuses on eliminating glucose- and galactose-containing foods and substituting fructose-containing foods. Fructose is absorbed by the brush border transporter GLUT5 by facilitated diffusion and is not dependent on *SLC5A1*.

Abetalipoproteinemia Abetalipoproteinemia is a rare disorder of lipid metabolism associated with abnormal erythrocytes (acanthocytes), neurologic symptoms, and steatorrhea (see Chap. 407). Lipolysis, micelle formation, and lipid uptake are all normal in patients with abetalipoproteinemia, but the re-esterified triglyceride cannot exit the epithelial cell because of the failure to produce chylomicrons. This disorder results from mutation of microsomal triglyceride transfer protein, which catalyzes the transfer of triglyceride onto nascent apolipoprotein B containing particles. Mutations in MTP decrease this transfer and decrease formation of chylomicrons. Small-intestinal biopsy samples obtained from these rare patients in the postprandial state reveal lipid-laden small-intestinal epithelial cells that become normal in appearance after a 72- to 96-h fast.

■ INTESTINAL MUCOSAL DISORDERS THAT RESULT IN MALABSORPTION OF MULTIPLE NUTRIENTS

Celiac Disease Celiac disease, also known as celiac sprue or gluten-sensitive enteropathy, is a small intestinal enteropathy that results from an immune response to gluten ingestion and is characterized by autoantibodies to tissue transglutaminase. Gluten is found in foods produced from wheat, rye, barley, and some varieties of oats, and it is a common additive to prepared foods and pharmaceuticals. Tissue transglutaminase is involved in the pathogenesis of this disorder, as it deamidates glutamine residues of gluten-derived peptides, facilitating their presentation by antigen-presenting cells.

Epidemiology and Genetics The incidence and prevalence of celiac disease have been increasing worldwide. Increased awareness among clinicians and patients has led to increases in detection, but there is evidence that the true incidence appears to be increasing as well. Global prevalence has been measured at 1.4%. In the United States, data from the National Health and Nutrition Examination survey showed seroprevalence of 0.2% in non-Hispanic black populations, 0.3% in Hispanic individuals, and 1.0% in white populations.

The prevalence of celiac disease is 10–15% in first-degree relatives. Host genetic factors include histocompatibility locus antigens HLADQ2 and DQ8; the presence of one of the two haplotypes is necessary but not sufficient for developing celiac disease. HLADQ2 and DQ8 are found in 25–35% of the general population; because most carriers never develop celiac disease, detection of these alleles is not useful for diagnosis. However, a negative test is very useful for ruling out celiac disease, with a negative predictive value of >99%. This is particularly helpful in patients who self-discontinued gluten ingestion prior to serologic or endoscopic testing.

Presentation Patients with celiac disease have a wide variety of disease manifestations, ranging from being asymptomatic, to having isolated iron-deficiency anemia due to duodenal disease, to severe diarrhea, weight loss, and malabsorption of multiple nutrients with more diffuse disease. Celiac disease primarily affects the proximal small intestine; it may involve the duodenum only or may cause widespread jejunal disease resulting in severe symptoms.

Diarrhea, weight loss, and growth failure in children are common presenting complaints, but additional signs and symptoms have become increasingly recognized to be associated with celiac disease, including bloating and irregular bowel habits, migraine headaches, and ataxia. In addition, patients may be identified after presenting with osteoporosis, iron-deficiency anemia, or detection of abnormal liver enzymes.

Mechanism of Diarrhea Patients with celiac disease have villus atrophy in the proximal small intestine and thus develop steatorrhea from mucosal malabsorption and may have lactase deficiency. However, they also develop a secretory component due to crypt hyperplasia and fluid hypersecretion from the crypt epithelium.

Associated Diseases Patients with celiac disease have a higher incidence of other autoimmune disorders such as type 1 diabetes mellitus and autoimmune thyroid disease. Dermatitis herpetiformis is a skin disorder that is highly associated with celiac disease, characterized by a vesicular rash mediated by IgA deposits in the skin. Down syndrome and Turner syndrome patients also have an increased risk of celiac disease.

Diagnosis Patients are screened for celiac disease first by testing for serum antibodies, including tissue transglutaminase IgA, anti-endomysial, and deamidated anti-gliadin antibodies. Serum IgA levels are measured to detect false-negative results from IgA deficiency. Deamidated anti-gliadin IgG antibodies or tissue transglutaminase IgG antibodies are detectable and diagnostic in IgA-deficient patients. The diagnosis in adults with positive antibody levels is confirmed by endoscopy with small-intestinal biopsy. Biopsies typically show characteristic villus blunting, crypt hyperplasia, and inflammation, including increased intraepithelial lymphocytes. The Marsh classification categorizes different types of celiac disease-related lesions and is currently used to quantify severity of disease involvement.

Family members of patients with celiac disease are screened if symptomatic; recommendations regarding screening asymptomatic family members are still controversial.

Complications Complications of celiac disease include refractory celiac disease, enteropathy-associated T-cell lymphoma, hyposplenism, and small-bowel adenocarcinoma.

Refractory Celiac Disease This complication is most common in patients with ongoing active celiac disease, found in about 10% of patients with persistent active disease. Patients have ongoing diarrhea and weight loss with persistent villus atrophy on biopsy after 1 year of following a strict gluten-free diet. These patients also have negative celiac serology, confirming their adherence to the gluten-free diet. Type 1 refractory celiac disease has a normal intraepithelial lymphocyte population whereas type 2 disease has clonal expansion of CD3+ intraepithelial lymphocytes that also contain a monoclonal rearrangement of the gamma chain of the T-cell receptor. Type 2 refractory celiac disease has a worse prognosis due to its association with T-cell lymphoma, which occurs in 33–50% of cases after 5 years. The therapy for celiac disease-related lymphoma is intense and includes high-dose chemotherapy and sometimes stem cell transplantation.

Small-bowel adenocarcinoma is a very rare cancer in the general population but is increased in celiac disease patients.

Therapy and Follow-up The mainstay of celiac disease treatment is institution of a strict gluten-free diet. This is challenging for patients because of the widespread presence of gluten in both raw and prepared foods, inaccurate food labeling, and cross-contamination during food preparation. Patients must receive rigorous dietary instruction from a dietitian and adhere lifelong to a gluten-free diet.

For those patients whose symptoms resolve, serologic follow-up is generally recommended to confirm compliance with a gluten-free diet. A follow-up biopsy to document complete healing of villus atrophy is also generally recommended. However, subsequent biopsies are not recommended unless symptoms recur. For patients without symptom resolution, a biopsy is required to determine the degree of disease activity and to rule out other causes of persistent diarrhea and complications such as refractory celiac disease or T-cell lymphoma. The most common cause of residual disease activity is dietary nonadherence or inadvertent gluten exposure. These patients pursue repeat consultation with a dietitian and efforts to reduce restaurant or other out-of-the-home exposure or cross-contamination at home. If biopsies are negative but symptoms persist, other causes of abdominal pain and diarrhea that are associated with celiac disease are considered, including irritable bowel syndrome, microscopic colitis, small-bowel bacterial overgrowth, and lactose or fructose intolerance.

Nonceliac Gluten Sensitivity Recently a subset of patients has been described with symptoms consistent with celiac disease but with negative serology and negative biopsies. Upon discontinuation of gluten, they have relief of abdominal pain, diarrhea, headaches/migraines, and other celiac disease-type symptoms. The etiology of this disorder is unknown.

■ WHIPPLE'S DISEASE

Whipple's disease is a chronic, multiorgan disease caused by *Tropheryma whipplei*, a gram-positive non-acid-fast, periodic acid-Schiff (PAS) positive rod, which is ubiquitous in the environment. Whipple's disease most commonly occurs in middle-aged men. Classic Whipple's disease is defined by the presence of arthralgias, weight loss, diarrhea, and abdominal pain. Other manifestations including central nervous system (CNS) and cardiac involvement are common and occur later in the disease. *T. whipplei* can be detected by polymerase chain reaction on involved tissue and is difficult to detect in the bloodstream. The intestinal lesion is also characterized by PAS-positive macrophages.

Clinical Presentation Arthralgias and arthritis are present for an average of 6 years before the GI symptoms begin, consistent with a persistent and substantial lag in diagnosis, which is still a problem today. Joint disease is present in >80% of patients. GI manifestations include

diarrhea, abdominal pain, and weight loss from malabsorption. CNS involvement is common and may include symptoms such as psychiatric manifestations or memory problems. Dementia and encephalitis may occur in later stages. Cardiac involvement may include endocarditis, pericarditis, and myocarditis.

Diagnosis For patients with GI manifestations, endoscopy with biopsies is performed and tissue is tested for *T. whipplei* by polymerase chain reaction. Tissue is also stained for PAS-positive macrophages and immunohistochemistry may also be performed to detect *T. whipplei*.

Treatment Prolonged antibiotics are recommended although the optimal regimen is still uncertain. Relapses are common, and often associated with the first manifestations of CNS involvement.

■ TROPICAL SPRUE

Tropical sprue is a poorly understood syndrome that is manifested by chronic diarrhea, steatorrhea, weight loss, and nutritional deficiencies, including both folate and vitamin B₁₂. Malabsorption of two unrelated substances is required for diagnosis. This disease occurs in 8–20% of people who have had an attack of infectious gastroenteritis in India, and is considered by some to be a postinfectious complication. It is prevalent in some but not all tropical areas, including southern India, Pakistan, the Philippines, Puerto Rico, Haiti, and Cuba. It occurs in residents as well as visitors to these areas.

Chronic diarrhea in a tropical environment is most often caused by infectious agents, including *Giardia lamblia*, *Yersinia enterocolitica*, *Entamoeba histolytica*, *C. difficile*, *Cryptosporidium parvum*, *Isospora belli*, *Strongyloides stercoralis*, and *Cyclospora cayetanensis*. Tropical sprue should not be entertained as a possible diagnosis until the presence of cysts and trophozoites has been excluded in three stool samples. **Chronic infections of the GI tract and diarrhea are discussed in Chaps. 46, 133, 134, 163–168, and 223.**

In the past few years, the term *environmental enteropathy* has been introduced as the diagnosis of many patients (especially infants and children) who had previously been diagnosed as tropical sprue. However, exact delineation of this newly designated entity is lacking.

Etiology Because tropical sprue responds to antibiotics, the consensus is that it may be caused by one or more infectious agents. Nonetheless, the etiology and pathogenesis of tropical sprue are uncertain. First, its occurrence is not evenly distributed in all tropical areas; it is rarely observed in Africa, Jamaica, or Southeast Asia. Second, an occasional individual does not develop symptoms of tropical sprue until long after having left an endemic area. For this reason, celiac disease (often referred to as celiac sprue) was originally called *nontropical sprue* to distinguish it from tropical sprue. Third, multiple microorganisms have been identified in jejunal aspirates, with relatively little consistency among studies. *Klebsiella pneumoniae*, *Enterobacter cloacae*, and *E. coli* have been implicated in some studies of tropical sprue, while other studies have favored a role for a toxin produced by one or more of these bacteria. Fourth, the incidence of tropical sprue appears to have decreased substantially during the past two or three decades, perhaps in relation to improved sanitation in many tropical countries during this time. Some have speculated that the reduced occurrence is attributable to the wider use of antibiotics in acute diarrhea, especially in travelers to tropical areas from temperate countries. Fifth, the role of folic acid deficiency in the pathogenesis of tropical sprue requires clarification. Folic acid is absorbed exclusively in the duodenum and proximal jejunum, and most patients with tropical sprue have evidence of folate malabsorption and depletion. Although folate deficiency can cause changes in small-intestinal mucosa that are corrected by folate replacement, several earlier studies reporting that tropical sprue could be cured by folic acid did not provide an explanation for the “insult” that was initially responsible for folate malabsorption.

The clinical pattern of tropical sprue varies in different areas of the world (e.g., India vs Puerto Rico). Not infrequently, individuals in southern India initially report the occurrence of acute enteritis before the development of steatorrhea and malabsorption. In contrast, in Puerto Rico, a more insidious onset of symptoms and a more dramatic

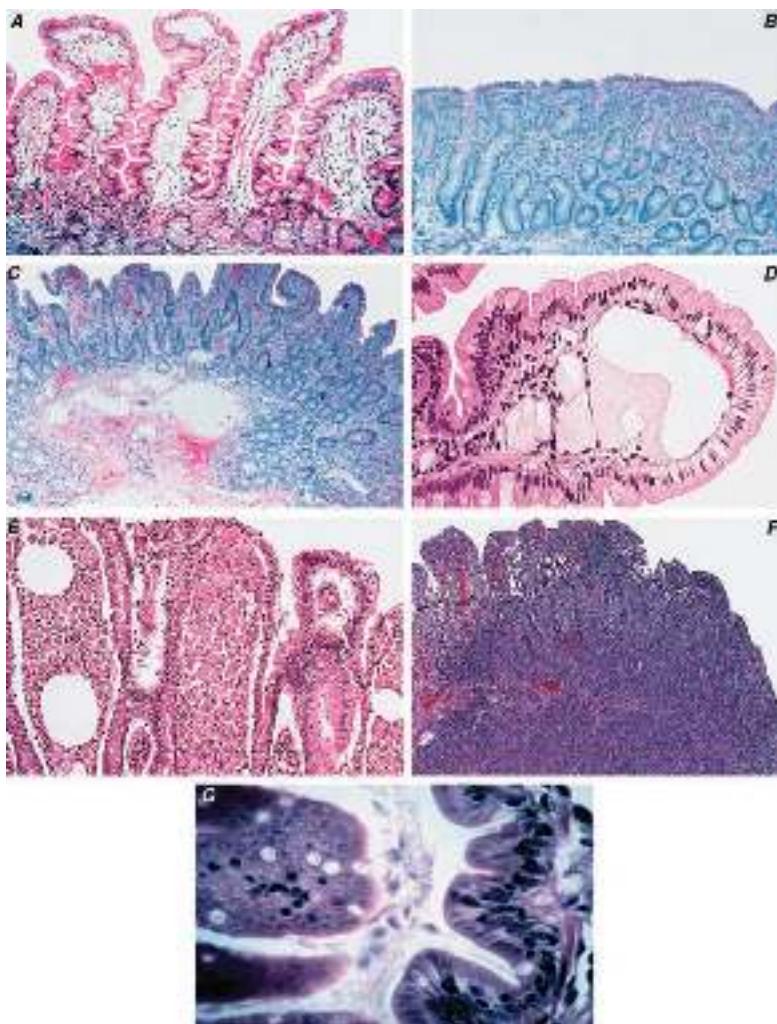


FIGURE 325-4 Small-intestinal mucosal biopsies. *A*, Normal individual. *B*, Untreated celiac disease. *C*, Treated celiac disease. *D*, Intestinal lymphangiectasia. *E*, Whipple's disease. *F*, Lymphoma. *G*, Giardiasis. (Courtesy of Marie Robert, MD, Yale University; with permission.)

response to antibiotics are seen compared with some other locations. Tropical sprue in different areas of the world may not be the same disease, and similar clinical entities may have different etiologies.

Diagnosis The diagnosis of tropical sprue is based on an abnormal small-intestinal mucosal biopsy in an individual with chronic diarrhea and evidence of malabsorption who is either residing or has recently lived in a tropical country. The small-intestinal biopsy in tropical sprue does not reveal pathognomonic features but resembles, and can often be indistinguishable from, that seen in celiac disease (Fig. 325-4). The biopsy sample in tropical sprue has less villous architectural alteration and more mononuclear cell infiltrate in the lamina propria. In contrast to those of celiac disease, the histologic features of tropical sprue manifest with a similar degree of severity throughout the small intestine, and a gluten-free diet does not result in either clinical or histologic improvement in tropical sprue.

TREATMENT

Tropical Sprue

Broad-spectrum antibiotics and folic acid are most often curative, especially if the patient leaves the tropical area and does not return. Tetracycline should be used for up to 6 months and may be associated with improvement within 1–2 weeks. Folic acid alone induces hematologic remission as well as improvement in appetite, weight

gain, and some morphologic changes in small-intestinal biopsy. Because of marked folate deficiency, folic acid is most often given together with antibiotics.

SHORT BOWEL SYNDROME

■ OVERVIEW

Short-bowel syndrome results from intestinal resection to treat a multitude of disorders including Crohn's disease, vascular diseases such as mesenteric arterial or venous thrombosis resulting in intestinal ischemia, volvulus, trauma, internal herniation, radiation enteritis, and diffuse carcinoma, among others. In children, the most common causes of short-bowel syndrome are necrotizing enterocolitis, intestinal atresias, volvulus, and malrotation. Short-bowel syndrome is defined as extensive removal of small intestine resulting in <200 cm remaining small bowel. Intestinal failure is functionally defined as persistent parenteral nutrition dependence, generally found in patients who have <100 cm of remaining small bowel and no residual colon in continuity.

Clinical Features Loss of small-bowel surface area in short-bowel syndrome results in severe diarrhea, weight loss, and malabsorption of multiple nutrients, including fat, protein, and carbohydrate. The severity of symptoms and ultimate dependence on parenteral nutrition are generally related to the extent of resection, presence or absence of residual colon in continuity, retention of the ileocecal valve, and

severity of the underlying disease. The intestine has a remarkable capacity to adapt to loss of small-bowel surface area, but this adaptive process is variable from patient to patient. Following resection, the adapting residual intestine exhibits an increase in crypt cell proliferation resulting in epithelial hyperplasia. The adaptive process generally continues for up to 2 years post resection, but improvements in nutrient, fluid, and electrolyte absorptive capacity have been reported even as late as 3–5 years after surgery. Massive diarrhea generally occurs in the first three postoperative months, associated with increased gastric acid secretion and malabsorption. Gradually, patients show enhanced functional capacity and reduced diarrhea. Specific nutrient deficiencies are dependent upon which segment of gut has been removed. For example, resection of the ileum results in loss of B_{12} absorptive and bile salt reabsorptive capacity. Malabsorbed bile salts reach the colon and cause a secretory diarrhea. In addition, resection of >100 cm of ileum results in such severe bile salt malabsorption that the liver cannot compensate by increased synthesis, thus precipitating fat malabsorption due to bile salt insufficiency/deficiency. Substantial resection of the colon also results in fluid and electrolyte loss and imbalance. The colon also plays a role in nutrient absorption because it metabolizes malabsorbed carbohydrate into short-chain fatty acids that can be absorbed by the colon and can contribute several hundred additional calories per day.

Long-Term Complications Because massive resection often leads to severe fat malabsorption, fat-soluble vitamin deficiency is common, and vitamin D deficiency can be very difficult to treat even with high-dose oral vitamin D supplementation, resulting in an increased risk of osteoporosis. Patients with a history of multiple surgeries often have extensive adhesive disease, and the residual intestine may have markedly abnormal motility or areas of structuring and narrowing, resulting in recurrent bacterial overgrowth. The frequency of renal calcium oxalate stones increases in patients with a shortened small bowel with an intact colon in continuity; calcium is saponified in the intestinal luminal contents that contain fatty acids, freeing oxalate to be absorbed in the colon resulting in hyperoxaluria.

Treatment The major focus of treatment for short-bowel syndrome is to control diarrhea and normalize nutrient, fluid, and electrolyte absorption so that patients can maintain their weight and have a healthy nutritional status without the support of parenteral nutrition. Medications include opiates and derivatives including loperamide and diphenoxylate-atropine, which slow intestinal motility to allow for more contact time between luminal nutrients and the small-bowel mucosal surface. In the first year following resection, acid-blocking medications are used to treat gastric hypersecretion, including proton pump inhibitors or histamine 2 antagonists. Small-bowel bacterial overgrowth is common and is treated with antibiotics if suspected. The only medication that is specific for short-bowel syndrome but limited for use in parenteral nutrition or intravenous fluid-dependent patients is teduglutide, a glucagon-like 2 peptide analog that enhances crypt cell proliferation and villus hyperplasia, and increases nutrient and fluid and electrolyte absorption. Patients treated with teduglutide have an average reduction of 20% of their parenteral nutrition requirements. Greater efficacy has been noted for patients without a residual colon, likely due to lower circulating endogenous GLP-2 levels compared to those with a colon in continuity.

Dietary Therapy Patients with short-bowel syndrome must consume three to four times their normal caloric intake to maintain their weight. The presence of luminal nutrients is required for the adaptive process to occur, so early feeding is recommended, even if parenteral nutrition is also required. These effects are most likely mediated by direct contact with the mucosa as well as stimulation of secretion of gut hormones such as glucagon-like 2.

If the patient has all or part of their colon remaining in continuity, a low-fat diet is instituted to reduce the concentration of malabsorbed fatty acids that induce a secretory diarrhea. High complex carbohydrates are encouraged because when malabsorbed and present in the colon, they are converted to short-chain fatty acids and are absorbed,

contributing several hundred additional kilocalories per day. All patients are asked to take a high-potency multivitamin on a daily basis. Patients in whom oral nutrition fails are fed with parenteral nutrition.

Monitoring SBS patients are at high risk for osteoporosis due to dietary calcium and vitamin D malabsorption, so they are periodically monitored for vitamin D deficiency and calcium levels and with DEXA studies to assess bone density. Malabsorption of vitamins and minerals is common; therefore, fat-soluble vitamins, vitamin B_{12} , folic acid, iron, magnesium, and zinc are monitored periodically. More unusual deficiencies include copper, selenium and chromium, but these are usually in PN-dependent patients and can be corrected by adjusting daily intravenous dosages. Signs and symptoms of vitamin and mineral deficiency are also carefully monitored (hair loss, skin and nail changes, neurologic symptoms such as peripheral neuropathy, etc.).

■ DISORDERS OF POST MUCOSAL ABSORPTION

Following uptake into enterocytes, nutrients are further processed and transported into the lymphatics or into the portal circulation for use by other cells throughout the body. Primary or secondary disorders of the lymphatics may result in significant diarrhea and malabsorption. Primary disorders of the intestinal lymphatics include intestinal lymphangiectasia, which may be congenital or acquired. Secondary causes of intestinal lymphatic damage or blockage include retroperitoneal fibrosis, fibrosing mesenteritis, and lymphoma. Circulatory causes of impaired delivery of nutrients from the intestine include Fontan physiology, congestive heart failure, and constrictive pericarditis. The end result of damage to lymphatic channels is malabsorption and diarrhea with concomitant protein-losing enteropathy.

PROTEIN LOSING ENTEROPATHY

Protein-losing enteropathy refers to a large group of GI and non-GI disorders characterized by hypoproteinemia and edema in the absence of liver disease with reduced protein synthesis, or kidney disease with proteinuria. These diseases are characterized by excess protein loss in the GI tract. Diseases that may result in increased protein loss into the GI tract can be classified into three groups: (1) mucosal ulceration, such that the protein loss primarily represents exudation across damaged mucosa (e.g., ulcerative colitis, GI carcinomas); (2) nonulcerated mucosa, but with evidence of mucosal damage so that the protein loss represents loss across epithelia with altered permeability (e.g., celiac disease and Ménétrier's disease [hypertrophic gastropathy] in the small intestine and stomach, respectively); and (3) lymphatic dysfunction, representing either primary lymphatic disease or lymphatic disease secondary to partial lymphatic obstruction that may occur as a result of enlarged lymph nodes or cardiac disease. These result in increased lymphatic pressure causing exudation of protein into the GI tract lumen.

Diagnosis The diagnosis of protein-losing enteropathy is suggested by diarrhea, peripheral edema, and low serum albumin and globulin levels in the absence of renal and hepatic disease. An individual with protein-losing enteropathy rarely has selective loss of *only* albumin or *only* globulins. Therefore, marked reduction of serum albumin with normal serum globulins should suggest renal and/or hepatic disease. Likewise, reduced serum globulins with normal serum albumin levels are more likely a result of reduced globulin synthesis rather than enhanced globulin loss into the intestine. Alpha-1 antitrypsin, a protein that accounts for ~4% of total serum proteins and is resistant to proteolysis, can be used to detect enhanced rates of serum protein loss into the intestinal tract but cannot be used to assess gastric protein loss because of its degradation in an acid milieu. Alpha-1 antitrypsin can be measured in a spot or 24-h stool collection and if elevated, is diagnostic. A more accurate determination is alpha-1 antitrypsin clearance, measured by determining stool volume as well as both stool and plasma alpha-1 antitrypsin concentrations. In addition to the loss of protein via abnormal and distended lymphatics, peripheral lymphocytes may be lost via lymphatics, with consequent relative lymphopenia and specifically loss of CD3+ T cells. Thus, lymphopenia in a patient with hypoproteinemia indicates increased loss of protein into the GI tract.

Patients with increased protein loss into the GI tract from lymphatic obstruction often have steatorrhea and diarrhea. The steatorrhea is a result of altered lymphatic flow as lipid-containing chylomicrons exit from intestinal epithelial cells via intestinal lymphatics (Table 325-2; Fig. 325-4). In the absence of mechanical or anatomic lymphatic obstruction, intrinsic intestinal lymphatic dysfunction—with or without lymphatic dysfunction in the peripheral extremities—has been designated *intestinal lymphangiectasia*. Similarly, ~50% of individuals with intrinsic peripheral lymphatic disease (Milroy's disease) also have intestinal lymphangiectasia and hypoproteinemia. Other than steatorrhea and enhanced protein loss into the GI tract, all other aspects of intestinal absorptive function are normal in intestinal lymphangiectasia.

Endoscopy and Imaging Endoscopy with biopsy and video capsule endoscopy may be performed to rule out mucosal disease. Magnetic resonance enterography may be helpful in children with lymphangiectasia.

Other Causes Patients who have idiopathic protein-losing enteropathy without evidence of GI disease should be examined for cardiac disease. As more patients with congenital heart disease reach adulthood, Fontan physiology has become a more common cause of protein-losing enteropathy. Other cardiac causes include right-sided valvular disease and chronic pericarditis (Chaps. 268 and 270). Ménétrier's disease (also called *hypertrophic gastropathy*) is an uncommon entity that involves the body and fundus of the stomach and is characterized by large gastric folds, reduced gastric acid secretion, and, at times, enhanced protein loss into the stomach.

TREATMENT

Protein-Losing Enteropathy

As excess protein loss into the GI tract is most often secondary to a specific disease, treatment should be directed primarily to the underlying disease process and not to the hypoproteinemia. When enhanced protein loss is secondary to lymphatic obstruction, it is critical to establish the nature of this obstruction. Identification of mesenteric nodes or lymphoma may be possible by imaging studies. Similarly, it is important to exclude cardiac disease as a cause of protein-losing enteropathy. Patients with congenital heart disease may be examined by intranodal lymphangiography or noncontrast magnetic resonance lymphangiography, and may undergo surgical lymphatic interventions to decompress the lymphatic system or to target exclusion of abnormal lymphatic channels.

The increased protein loss that occurs in intestinal lymphangiectasia is a result of distended lymphatics associated with lipid malabsorption. The hypoproteinemia is treated with a low-fat, high-protein diet and the administration of medium-chain triglycerides, which do not exit from the intestinal epithelial cells via lymphatics but are delivered to the body via the portal vein. Other medical therapies including octreotide, a somatostatin analog, intravenous heparin, and budesonide have been studied but have generally been ineffective.

APPROACH TO THE PATIENT

Evaluation of the Patient with Suspected Malabsorption

The evaluation of patients with malabsorption is often challenging due to the large number of underlying disorders and the wide array of available tests. Thus an extensive history and careful physical examination are essential to develop a more limited differential diagnosis, and thereby avoid extensive and unnecessary testing.

HISTORY

A careful history should include questions about symptoms including abdominal pain, diarrhea, weight loss, bloating,

symptoms or signs of selective nutrient deficiency including iron deficiency anemia, bone fracture, or osteoporosis suggesting vitamin D and/or calcium deficiency, peripheral neuropathy resulting from vitamin B₁₂ deficiency, hair loss that may result from generalized protein deficiency, predisposing disorders such as chronic pancreatitis or liver disease particularly involving the bile ducts such as primary biliary cholangitis or primary sclerosing cholangitis, history of small-bowel resection (due to Crohn's disease, trauma, ischemic bowel disease, etc.), and travel history. A multitude of nonspecific symptoms such as fatigue and weakness may also be reported. The protean manifestations of malabsorption and the underlying pathophysiology of clinical manifestations are summarized in Table 325-5.

PHYSICAL EXAMINATION

A careful physical examination may provide clues to underlying nutrient deficiencies and help assess severity of the malabsorptive process. For example, evidence of significant weight loss may be detected by bitemporal wasting and reduced arm circumference, iron deficiency may cause nail spooning, and vitamin B₁₂ deficiency may result in significant peripheral neuropathy resulting in sensory reduction with tingling or numbness.

LABORATORY EXAMINATION (TABLE 325-6)

Diseases that exclusively affect the proximal small intestine (e.g., celiac disease limited to the duodenum) may result in iron-deficiency anemia. Resection or disease of the terminal ileum frequently results in B₁₂ deficiency since B₁₂ absorption occurs exclusively in the ileum, causing a macrocytic anemia. Disorders that cause steatorrhea are almost invariably associated with fat-soluble vitamin deficiency, specifically vitamin D (very common), vitamin E, vitamin A, and vitamin K. The functional result of vitamin K deficiency is an elevated prothrombin time/international normalized ratio (INR) so this blood test is frequently measured instead of vitamin K levels. Serum carotene levels can suggest fat malabsorption but may decrease simply due to poor dietary consumption of leafy vegetables.

To diagnose steatorrhea, a spot stool can be submitted for Sudan III staining, which is specific for fecal fat. This is a useful qualitative but not quantitative test. Stool for elastase is helpful for diagnosing pancreatic insufficiency. A 24-h assessment of stool volume/weight may be useful to establish the presence of clinically significant absorptive or secretory diarrhea vs diarrhea from other causes such as proctitis, which causes frequent, small, low-volume stools. The gold standard for documenting steatorrhea is the 72-h fecal fat collection, which is performed in concert with the patient's consumption of a 100-g fat diet. This test is highly accurate but difficult to obtain due to patient reluctance to collect stool. Also patients with fat malabsorption may poorly tolerate a 100-g fat diet. A diet with strictly quantified albeit reduced fat calories may be substituted. Finally, the calculation of the stool osmotic gap is a very useful and easy way to diagnose an osmotic diarrhea. A spot stool sample is sent to the lab for quantitation of fecal sodium and potassium concentration. Although stool osmolality can also be measured in the lab, measurements are often inaccurate due to bacterial degradation of nonabsorbed carbohydrate as the stool sits prior to examination. Because normal stool osmolality reflects serum osmolality at 290 mOsm/kg H₂O, the osmotic gap may be calculated as follows:

$$290 - 2 \text{ (stool [Na+] + stool [K+])}$$

If >50–100, a stool osmotic gap is present indicating the presence of unmeasured osmoles (e.g., malabsorbed lactose), and osmotic diarrhea can be diagnosed. If <50, one can presume a secretory component. Of note, malabsorbed fatty acids may also cause a secretory diarrhea by inducing secretion in the colon, so a malabsorptive diarrhea may have both an osmotic and secretory component. Extensive celiac disease may cause both osmotic diarrhea due to malabsorbed carbohydrate and also secretory diarrhea due to crypt hyperplasia.

TABLE 325-5 Pathophysiology of Clinical Manifestations of Malabsorption Disorders

SYMPOTM OR SIGN	MECHANISM
Weight loss/malnutrition	Anorexia, malabsorption of nutrients
Diarrhea	Impaired absorption or secretion of water and electrolytes; colonic fluid secretion secondary to unabsorbed dihydroxy bile acids and fatty acids
Ratus	Bacterial fermentation of unabsorbed carbohydrate
Glossitis, cheilosis, stomatitis	Deficiency of iron, vitamin B ₁₂ , folate, and vitamin A
Abdominal pain	Bowel distention or inflammation, pancreatitis
Bone pain	Calcium, vitamin D malabsorption, protein deficiency, osteoporosis
Tetany, paresthesia	Calcium and magnesium malabsorption
Weakness	Anemia, electrolyte depletion (particularly K ⁺)
Azotemia, hypotension	Fluid and electrolyte depletion
Amenorrhea, decreased libido	Protein depletion, decreased calories, secondary hypopituitarism
Anemia	Impaired absorption of iron, folate, vitamin B ₁₂
Bleeding	Vitamin K malabsorption, hypoprothrombinemia
Night blindness/xerophthalmia	Vitamin A malabsorption
Peripheral neuropathy	Vitamin B ₁₂ and thiamine deficiency
Dermatitis	Deficiency of vitamin A, zinc, and essential fatty acid

Urinary D-xylose Test The urinary D-xylose test for carbohydrate absorption provides a measure of proximal small-bowel absorptive function. -Xylose, a pentose, is absorbed almost exclusively in the proximal small intestine and is excreted in the urine. The -xylose test is usually performed by administering 25 g of -xylose and collecting urine for 5 h. An abnormal test (excretion of <4.5 g) primarily reflects duodenal/jejunal mucosal disease. The -xylose test can also be abnormal in patients with delayed gastric emptying, impaired renal function, and sequestration in patients with large collections of fluid in a third space (i.e., ascites, pleural fluid). The ease of obtaining a mucosal biopsy of the small intestine by endoscopy and the false-negative rate of the -xylose test have led to its diminished use. When small-intestinal mucosal disease is suspected, a small-intestinal mucosal biopsy should be performed.

Radiologic Examination A small-bowel follow-through barium examination may be very useful for detecting evidence of small-bowel diseases such as celiac disease, jejunal diverticulosis that predisposes to small-bowel bacterial overgrowth, or Crohn's disease (Fig. 325-3). Magnetic resonance enterography and CT enterography are commonly used for diagnosis and management of inflammatory, stricturing disorders such as Crohn's disease and as an initial assessment of malabsorption, providing a means to visualize the entire luminal GI tract as well as the hepatobiliary tree and pancreas.

Endoscopic Evaluation and Small-Bowel Biopsies Endoscopy with small-bowel biopsy is essential in the evaluation of patients

with documented steatorrhea or chronic diarrhea, as well as to evaluate abnormalities detected by radiologic imaging or by capsule endoscopy. In patients with documented steatorrhea and no evidence of pancreatic or hepatobiliary disease, an upper endoscopy and possible small-bowel enteroscopy are required to examine the small-bowel mucosa and to take biopsies for analysis. An upper endoscopy will visualize the stomach and duodenum; the maximum reach of the typical upper endoscopy scope is the ligament of Treitz. Small-bowel enteroscopy using a longer scope such as a pediatric colonoscope can be used to visualize the jejunum. Single- and double-balloon enteroscopy provide a means for examining much more of the jejunum and, if successful, will reach the ileum. Capsule endoscopy provides another means for visualizing the entire small bowel. Colonoscopy can be used for a retrograde view and biopsy of the terminal ileum.

Biopsy Analysis Small-bowel pathology may be divided into the three groups (Table 325-7) described below.

1. Diffuse histopathologic findings involving the entire or majority of the mucosa which are specific for a particular disease entity; these include Whipple's disease, agammaglobulinemia (for example, combined variable immunodeficiency), and abetalipoproteinemia. Whipple's disease exhibits PAS-positive macrophages and immunohistochemical analysis can detect the pathogenic organism. Immune globulin deficiency is associated with a variety of histopathologic findings on small-intestinal mucosal biopsy. The characteristic feature is the absence of or substantial reduction in the number of plasma cells in the lamina propria; the mucosal architecture may be either perfectly normal or flat (i.e., villous atrophy). Abetalipoproteinemia is characterized by a normal mucosal appearance except for the presence of mucosal absorptive cells that contain lipid postprandially and disappear after a prolonged period of either fat-free intake or fasting.
2. Patchy lesions that are specific for a disease entity include, for example, intestinal lymphoma or intestinal lymphangiectasia. Several diseases feature an abnormal small-intestinal mucosa with a patchy distribution. As a result, biopsy samples obtained randomly or in the absence of endoscopically visualized abnormalities may not reveal diagnostic features. Intestinal lymphoma can at times be diagnosed on mucosal biopsy by the identification of malignant lymphoma cells in the lamina propria and submucosa (Chap. 108). Dilated lymphatics in the submucosa and sometimes in the lamina propria indicate lymphangiectasia associated with hypoproteinemia secondary to protein loss into the intestine. Eosinophilic gastroenteritis comprises a heterogeneous group of disorders with a spectrum of presentations and symptoms, with an eosinophilic infiltrate of the lamina propria, and with or without peripheral eosinophilia. The patchy nature of the infiltrate and its presence in the submucosa often lead to an absence of histopathologic findings on mucosal biopsy. As the involvement of the duodenum in Crohn's disease is also submucosal and not necessarily continuous, mucosal biopsies are not the most direct approach to the diagnosis of duodenal Crohn's disease (Chap. 326). Amyloid deposition can be identified by

TABLE 325-6 Comparison of Different Types of Fatty Acids

	LONG-CHAIN	MEDIUM-CHAIN	SHORT-CHAIN
Carbon chain length	>12	8–12	<8
Present in diet	In large amounts	In small amounts	No
Origin	In diet as triglycerides	Only in small amounts in diet as triglycerides	Bacterial degradation in colon of nonabsorbed carbohydrate to fatty acids
Primary site of absorption	Small intestine	Small intestine	Colon
Requires pancreatic lipolysis	Yes	No	No
Requires micelle formation	Yes	No	No
Present in stool	Minimal	No	Substantial

TABLE 325-7 Diseases That Can Be Diagnosed by Small-Intestinal Mucosal Biopsies

LESIONS	PATHOLOGIC FINDINGS
Diffuse, Specific	
Whipple's disease	Lamina propria includes macrophages containing material positive on periodic acid-Schiff staining
Agammaglobulinemia	No plasma cells; either normal or absent villi ("flat mucosa")
Abetalipoproteinemia	Normal villi; epithelial cells vacuolated with fat postprandially
Patchy, Specific	
Intestinal lymphoma	Malignant cells in lamina propria and submucosa
Intestinal lymphangiectasia	Dilated lymphatics; clubbed villi
Eosinophilic gastroenteritis	Eosinophil infiltration of lamina propria and mucosa
Amyloidosis	Amyloid deposits
Crohn's disease	Noncaseating granulomas
Infection by one or more microorganisms (see text)	Specific organisms
Mastocytosis	Mast cell infiltration of lamina propria
Diffuse, Nonspecific	
Celiac disease	Short or absent villi; mononuclear infiltrate; epithelial cell damage; hypertrophy of crypts
Tropical sprue	Similar to celiac disease
Bacterial overgrowth	Patchy damage to villi; lymphocyte infiltration
Folate deficiency	Short villi; decreased mitosis in crypts; megalocytosis
Vitamin B ₁₂ deficiency	Similar to folate deficiency
Radiation enteritis	Similar to folate deficiency
Zollinger-Ellison syndrome	Mucosal ulceration and erosion from acid
Protein-calorie malnutrition	Villous atrophy; secondary bacterial overgrowth
Drug-induced enteritis	Variable histology

Congo Red staining in some patients with amyloidosis involving the duodenum ([Chap. 112](#)).

3. Diffuse nonspecific lesions may be found in more than one disorder. For example, villus atrophy/absence may be found in celiac disease, tropical sprue, or bacterial overgrowth, among other disorders. Several microorganisms can be identified in small-intestinal biopsy samples, establishing a correct diagnosis. At times, the biopsy is performed specifically to diagnose infection (e.g., Whipple's disease or giardiasis). In most other instances, the infection is detected incidentally during the workup for diarrhea or other abdominal symptoms. Many of these infections occur in immunocompromised patients with diarrhea; the etiologic agents include *Cryptosporidium*, *Isospora belli*, microsporidia, *Cyclospora*, *Toxoplasma*, cytomegalovirus, adenovirus, *Mycobacterium avium-intracellulare*, and *G. lamblia*. In immunocompromised patients, when *Candida*, *Aspergillus*, *Cryptococcus*, or *Histoplasma* organisms are seen on duodenal biopsy, their presence generally reflects systemic infection. Apart from Whipple's disease and infections in the immunocompromised host, small-bowel biopsy is seldom used as the primary mode of diagnosis of infection. Even giardiasis is more easily diagnosed by stool antigen studies and/or duodenal aspiration than by duodenal biopsy.

SUMMARY

The evaluation and management of patients with disorders of absorption is challenging due to the complexity of the underlying pathophysiology and the large number of associated diseases. A diagnostic approach based on the information summarized in Tables 325-1 and

325-5 should prove useful for guiding the care of these challenging patients.

A

Henry Binder wrote this chapter in prior editions and some material from his chapter has been retained.

■ FURTHER READING

- B HJ, R DC: Short bowel syndrome, in *Gastrointestinal Motility Disorders: A Point-of-Care Clinical Guide*. E Bardan, R Shaker (eds). Cham, Switzerland, Springer International Publishing, 2017.
- B D, Q EM: Small intestinal bacterial overgrowth. *Gastroenterol Clin North Am* 50:463, 2021.
- C G et al: Celiac disease: A comprehensive current review. *BMC Med* 17:142, 2019.
- C M, V P: The role of bile acids in chronic diarrhea. *Am J Gastroenterol* 115:1596, 2020.
- E L et al: Protein-losing enteropathy. *Curr Opin Gastroenterol* 36:238, 2020.
- J LR: Digestion and absorption of nutrients, in *Gastrointestinal Physiology*, 9th ed. LR Johnson. Philadelphia, Elsevier, 2019, pp 102-120.
- L JC, R D: Whipple's disease and *Tropheryma whipplei* infections: When to suspect them and how to diagnose and treat them. *Curr Opin Infect Dis* 31:463, 2018.
- L B, R -T A: Epidemiology, presentation and diagnosis of celiac disease. *Gastroenterology* 160:63, 2021.
- L DG, L MD: Protein losing enteropathy: comprehensive review of the mechanistic association with clinical and subclinical disease states. *Clin Exper Gastro* 10:247, 2017.
- M B et al: Update on lactose malabsorption and intolerance: pathogenesis, diagnosis and clinical management. *Gut* 68: 2080, 2019.

326

Inflammatory Bowel Disease

Sonia Friedman, Richard S. Blumberg



Inflammatory bowel disease (IBD) is a chronic idiopathic inflammatory disease of the gastrointestinal tract. Ulcerative colitis (UC) and Crohn's disease (CD) are the two major types of IBD.

■ GLOBAL CONSIDERATIONS: EPIDEMIOLOGY

UC and CD have emerged as global diseases in the twenty-first century. They affect >2 million individuals in North America, 3.2 million in Europe, and millions more worldwide. Since the late 1990s, the majority of studies on CD and UC show stable or falling incidence in the Western world. The disease burden remains high, with a prevalence of >0.3% in North America, Oceania, and most countries in Europe. In newly industrialized countries in Africa, Asia, and South America where there is increased urbanization and Westernization, the incidence of IBD has been rising and mirrors the prior increase of IBD in the Western world in the twentieth century. For example, in Brazil, the annual percent change is +11.1% (95% confidence interval [CI], 4.8–17.8%) for CD and +14.9% (95% CI, 10.4–19.6%) for UC, whereas in Taiwan, the annual percent change is +4.0% (95% CI, 1.0–7.1%) for CD and +4.8% (95% CI, 1.8–8.0%) for UC. In a study of newly diagnosed IBD cases between 2011 and 2013 from 13 countries or regions in the Asia Pacific, the mean annual IBD incidence per 100,000 was 1.50 (95% CI, 1.43–1.57). India (9.31; 95% CI, 8.38–10.31) and China (3.64; 95% CI, 2.97–4.42) had the highest IBD incidences in Asia. The highest reported prevalence values were in Europe (UC, 505 per 100,000 in

TABLE 326-1 Epidemiology of IBD

	ULCERATIVE COLITIS	CROHN'S DISEASE
Age of onset	Second to fourth decades and seventh to ninth decades	Second to fourth decades and seventh to ninth decades
Ethnicity	Jewish > non-Jewish white > black > Latinx > Asian	
Female-to-male ratio	0.51–1.58	0.34–1.65
Smoking	May prevent disease (odds ratio 0.58)	May cause disease (odds ratio 1.76)
Oral contraceptives	No increased risk	Hazard ratio 2.82
Appendectomy	Protective (risk reduction 13–26%)	Not protective
Monozygotic twins	6–18% concordance	38–58% concordance
Dizygotic twins	0–2% concordance	4% concordance
Infections in the first year of life	1.6 and 3 times the risk of developing IBD by age 10 and 20 years	

Abbreviation: IBD, inflammatory bowel disease.

Norway; CD, 322 per 100,000 in Germany) and North America (UC, 286 per 100,000 in the United States; CD, 319 per 100,000 in Canada). The most likely factors that explain the geographic variability of IBD rates, especially the rising incidence in developing countries and urban areas, are environmental variables including changes in diet (with downstream effects on the intestinal microbiota), exposure to sunlight or temperature differences, and socioeconomic status and hygiene (Table 326-1).

Increasing immigration to Western societies also has an impact on the incidence and prevalence of IBD. The prevalence of UC among southern Asians who immigrated to the United Kingdom (UK) was higher in comparison to the European UK population (17 cases per 100,000 persons vs 7 per 100,000). Spanish patients who emigrated within Europe, but not those who immigrated to Latin America, developed IBD more frequently than controls. Individuals who have immigrated to Westernized countries and then returned to their country of birth also continue to demonstrate an increased risk of developing IBD.

Peak incidence of UC and CD is in the second to fourth decades, with 78% of CD studies and 51% of UC studies reporting the highest incidence among those aged 20–29 years old. A second modest rise in incidence occurs between the seventh and ninth decades of life. The female-to-male ratio ranges from 0.51 to 1.58 for UC studies and 0.34 to 1.65 for CD studies, suggesting that the diagnosis of IBD is not gender-specific. Pediatric IBD (patients <17 years old) composes ~20–25% of all IBD patients, and ~5% of all IBD patients are <10 years old. Children with IBD are also grouped as those with early-onset (EO) IBD (patients <10 years old), very-early-onset (VEO) IBD (patients <6 years old), and infantile IBD (patients <2 years old). VEOIBD and infantile IBD mainly affect the colon and are resistant to standard medications, and patients often have a strong family history of IBD, with at least one first-degree relative affected. In infantile IBD or VEOIBD, a number of rare, single genetic mutations have been identified as the basis for this susceptibility in up to 10% of patients, suggesting a simple Mendelian origin of the disease in these cases.

The greatest incidence of IBD is among white and Jewish people, but the incidence of IBD in Latinx and Asian people is increasing, as noted above. Urban areas have a higher prevalence of IBD than rural areas, and high socioeconomic classes have a higher prevalence than lower socioeconomic classes.

Epidemiologic studies have identified a number of potential environmental factors that are associated with disease risk (Fig. 326-1). In Caucasian populations, smoking is an important risk factor in IBD with opposite effects on UC (odds ratio [OR] 0.58) and CD (OR 1.76), whereas in other ethnic groups with different genetic susceptibility, smoking may play a lesser role. Previous appendectomy with confirmed appendicitis (risk reduction of 13–26%), particularly at a young age, has a protective effect on the development of UC across different geographical regions and populations. Appendectomy is modestly associated with the development of CD, but this may be due to

diagnostic bias. Oral contraceptive use is associated with an increased risk of CD, with a reported hazard ratio as high as 2.82 among current users and 1.39 among past users. The association between oral contraceptive use and UC is limited to women with a history of smoking. Infections in the first year of life are associated with development of IBD, especially before the ages of 10 and 20 years. Breast-feeding may also protect against the development of IBD. Infectious gastroenteritis with pathogens (e.g., *Salmonella*, *Shigella*, *Campylobacter* spp., *Clostridium difficile*) increases IBD risk by two- to threefold. Diets high in animal protein, sugars, sweets, oils, fish and shellfish, and dietary fat, especially ω-6 fatty acids, and low in ω-3 fatty acids have been implicated in increasing the risk of IBD. A protective effect of vitamin D on the risk of CD has been reported.

IBD is a familial disease in 5–10% of patients (Fig. 326-2), and the strongest risk factor for the development of IBD is a first-degree relative with the disease. The children of mothers and fathers with UC have an approximately fourfold increased risk of UC, and the children of mothers and fathers with CD have an almost eightfold increased risk of CD. Some of these patients may exhibit early-onset disease during the first decade of life and, in CD, a concordance of anatomic site and clinical type within families. In twin studies, 38–58% of monozygotic twins are concordant for CD, and 6–18% are concordant for UC, whereas 4% of dizygotic twins are concordant for CD, and 0–2% are concordant for UC in Swedish and Danish cohorts. In the remainder of patients, IBD is observed in the absence of a family history (i.e., sporadic disease).

GLOBAL CONSIDERATIONS: IBD PHENOTYPES

IBD location and behavior show racial differences that may reflect underlying genetic variations and have important implications for diagnosis and management of disease. Blacks and Latinxs tend to have an ileocolonic CD distribution. Data from East Asia show that ileocolonic CD is the most common CD phenotype (50.5–71%) and perianal disease is more common in East Asian patients (30.3–58.8%) than whites (25.1–29.6%). Pancolonic disease is more common than left-sided colitis or proctitis among black, Latinx, and Asian patients with UC. Older Asian patients with UC (age >60) tend to have a more aggressive disease course. Among blacks, joint involvement is the predominant extraintestinal manifestation (EIM) reported and ranges from 15.7 to 29.6%. Ocular involvement is also common in African Americans and ranges from 7.1 to 13%. Dermatologic manifestations are the most common EIMs reported in Latinxs (10–13%). Few data shed light on all aspects of disease in Hispanics, on the incidence and prevalence of IBD in blacks, and in Asians with IBD outside of Asia. These ethnic variations indicate the importance of different genetic and/or environmental factors in the pathogenesis of this disorder.

ETIOLOGY AND PATHOGENESIS

Under physiologic conditions, homeostasis normally exists between the commensal microbiota, epithelial cells that line the interior of the intestines (intestinal epithelial cells [IECs]), and immune cells within the tissues (Fig. 326-1). A consensus hypothesis is that each of these three major host compartments that function together as an integrated “supraorganism” (microbiota, IECs, and immune cells) are affected by specific environmental (e.g., smoking, antibiotics, enteropathogens) and genetic factors that, in a susceptible host, cumulatively and interactively disrupt homeostasis during the course of one's life and, in so doing, culminate in a chronic state of dysregulated inflammation; i.e., IBD. Although chronic activation of the mucosal immune system may represent an appropriate response to an infectious agent, a search for such an agent has thus far been unrewarding in IBD. As such, IBD is currently considered an inappropriate immune response to the endogenous (autochthonous) commensal microbiota within the intestines, with or without some component of autoimmunity. Importantly, the normal, uninflamed intestines contain a large number of immune cells that are in a unique state of activation, in which the gut is restrained from full immunologic responses to the commensal microbiota and dietary antigens by very powerful regulatory pathways that function within the immune system (e.g., T regulatory cells that express the

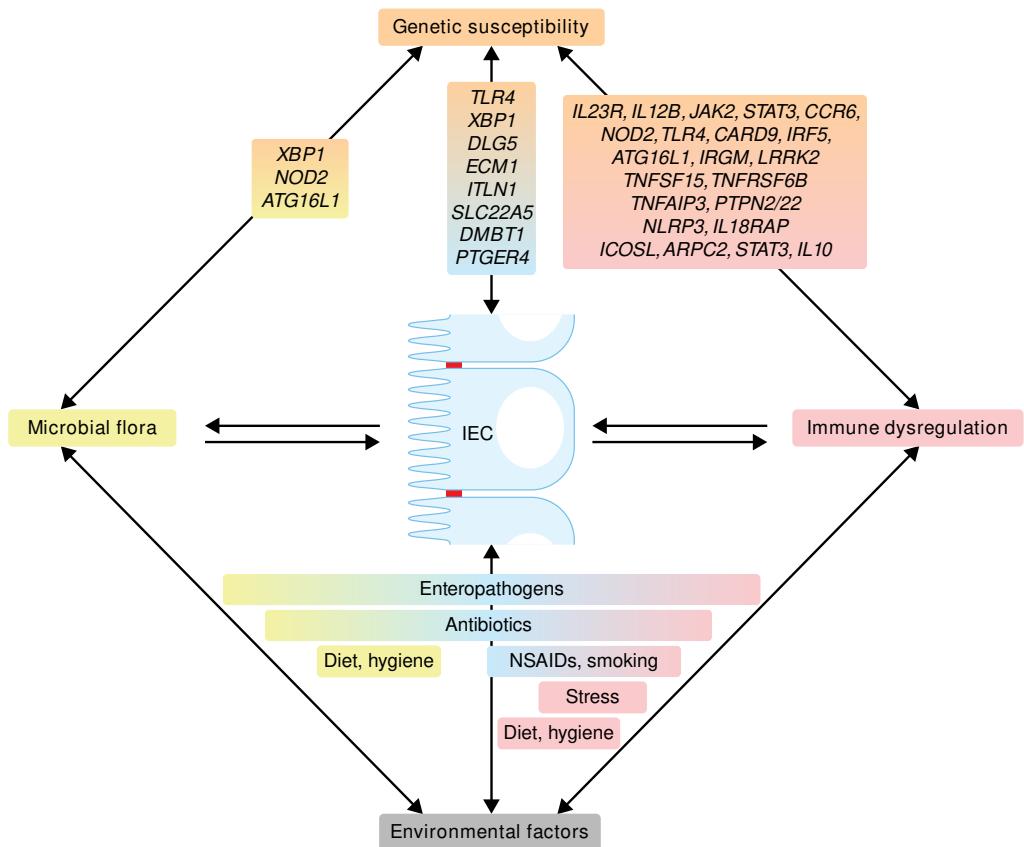


FIGURE 326-1 Pathogenesis of inflammatory bowel disease (IBD). In IBD, the tridirectional relationship between the commensal flora (microbiota), intestinal epithelial cells (IECs), and mucosal immune system is dysregulated, leading to chronic inflammation. Each of these three factors is affected by genetic and environmental factors that determine risk for the disease. NSAIDs, nonsteroidal anti-inflammatory drugs. (Republished with permission Annual Review of Immunology from Inflammatory Bowel Disease, A Kaser et al: 28:573, 2010. Permission conveyed through Copyright Clearance Center, Inc.)

FoxP3 transcription factor and suppress inflammation). Maintenance of homeostasis also involves oversight from local parenchymal cells including nerve, endothelial, and stromal cells, as well as the commensal microbiota that provide essential remedial factors necessary for health and serve as a target of the immune response. During the course of infections or other environmental stimuli in the normal host, full activation of the lymphoid tissues in the intestines occurs but is rapidly superseded by dampening of the immune response and tissue repair. In IBD, such processes may not be regulated normally.

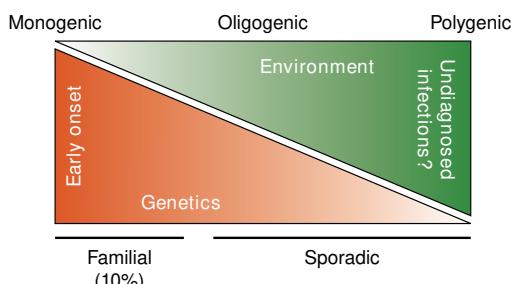


FIGURE 326-2 A model for the syndromic nature of inflammatory bowel disease (IBD). Genetic and environmental factors variably influence the development and phenotypic manifestations of IBD. At the one extreme, IBD is exemplified as a simple Mendelian disorder as observed in early-onset IBD due to single-gene defects such as *IL10*, *IL10RA*, and *IL10RB*; and at the other extreme, it may be exemplified by as yet to be described emerging infectious diseases. (Reproduced with permission from A Kaser et al: *Genes and environment: how will our concepts on the pathophysiology of IBD develop in the future?*, *Dig Dis* 28:395, 2010.)

GENETIC CONSIDERATIONS

The genetic underpinning of IBD is known from its concordance in identical twins, its occurrence in the context of several genetic syndromes, and the development of severe, refractory IBD in early life in association with single-gene defects that affect the immune system (Table 326-2). More than 60 different gene defects have been identified in patients with VEOIBD by whole exome sequencing (WES), in whom the majority of monogenic mutations have been discovered. These include mutations in genes encoding, for example, interleukin (IL) 10, the IL-10 receptor (IL-10R), cytotoxic T-lymphocyte-associated protein-4 (*CTLA4*), neutrophil cytosolic factor 2 protein (*NCF2*), X-linked inhibitor of apoptosis protein (*XIAP*), lipopolysaccharide responsive and beige-like anchor protein (*LRBA*), and tetratricopeptide repeat domain 7A protein (*TTC7*), among many other genes that are involved in host-commensal interactions. A monogenic etiology may also be possible in a small subset of adult patients with IBD. In addition, IBD has a familial origin in at least 10% of afflicted individuals, consistent with an inherited basis for this disease (Fig. 326-2). However, the majority of cases of pediatric (non-VEOIBD) and adult IBD are multigenic (or polygenic) in origin, suggesting a syndromic nature of this disease that gives rise to multiple clinical subgroups beyond the simple classification as UC and CD. The polygenic nature of the disease has been elucidated through a variety of genetic approaches, including candidate gene studies, linkage analysis, and genome-wide association studies (GWAS) that focus on the identification of disease-associated single nucleotide polymorphisms (SNPs) within the human genome and WES and whole genome sequencing to elucidate the specific mutations potentially involved. GWAS have identified ~240 genetic loci; two-thirds of these loci are associated with both disease phenotypes,

TABLE 326-2 Primary Genetic Disorders Associated with IBD

NAME	GENETIC ASSOCIATION	PHENOTYPE
Turner's syndrome	Loss of part or all of X chromosome	Associated with UC and colonic CD
Hermansky-Pudlak syndrome	Autosomal recessive chromosome 10q23	Granulomatous colitis, oculocutaneous albinism, platelet dysfunction, pulmonary fibrosis
Wiskott-Aldrich syndrome (WAS)	X-linked recessive disorder, loss of WAS protein function	Colitis, immunodeficiency, severely dysfunctional platelets, and thrombocytopenia
Glycogen storage disease type B1	Autosomal recessive disorder of <i>SLC37A4</i> resulting in deficiency of the glucose-6-phosphate translocase	Granulomatous colitis, presents in infancy with hypoglycemia, growth failure, hepatomegaly, and neutropenia
Immune dysregulation polyendocrinopathy, enteropathy X-linked (IPEX)	Loss of FoxP3 transcription factor and T regulatory cell function	UC-like autoimmune enteropathy, with endocrinopathy (neonatal type 1 diabetes or thyroiditis), dermatitis
Early-onset IBD	Deficient IL-10 and IL-10 receptor function	Severe, refractory IBD in early life

Abbreviations: CD, Crohn's disease; IBD, inflammatory bowel disease; IL, interleukin; UC, ulcerative colitis.

with the remainder being specific for either CD or UC (Table 326-3). These genetic similarities account for the overlapping immunopathogenesis and consequently epidemiologic observations of both diseases in the same families and similarities in response to therapies. Because

the specific causal variants for each identified gene or locus are mostly unknown as most risk loci are contained within regulatory (noncoding) regions of the associated genes, it is not clear whether the similarities in the genetic risk factors associated with CD and UC are shared at a structural or functional level. The risk conferred by each identified gene or locus is unequal and generally small, such that only ~20% of the disease risk is considered to be explained by the current genetic information. Further, many of the genetic risk factors identified are also observed to be associated with risk for other immune-mediated diseases, suggesting that related immunogenetic pathways are involved in the pathogenesis of multiple different disorders, accounting for the common responsiveness to similar types of biologic therapies (e.g., anti-tumor necrosis factor [TNF] therapies) and possibly the simultaneous occurrence of these disorders. The diseases and the genetic risk factors that are shared with IBD include, for example, rheumatoid arthritis (*TNFAIP3*), psoriasis (*IL23R*, *IL12B*), ankylosing spondylitis (*IL23R*), type 1 diabetes mellitus (*IL10*, *PTPN2*), asthma (*ORMDL3*), and systemic lupus erythematosus (*TNFAIP3*, *IL10*), among others.

The genetic factors that are recognized to mediate risk for IBD have highlighted the importance of shared mechanisms of disease that variably affect CD and/or UC (Table 326-3). These include the following: those genes that are associated with fundamental cell biologic processes such as the unfolded protein response due to endoplasmic reticulum stress, autophagy, and metabolism that regulate the ability of cells to manage the physiologic needs of the intestinal environment; those associated with innate immunity associated with nonlymphoid cells that function in responses to and control of microbes; those associated with the regulation of adaptive immunity that control the balance between inflammatory and anti-inflammatory cellular pathways associated with lymphocytes; and, finally, those that are involved in the development and resolution of inflammation associated with

TABLE 326-3 Some Genetic Loci Associated with Crohn's Disease and/or Ulcerative Colitis

CHROMOSOME	PUTATIVE GENE	GENE NAME	PROTEIN FUNCTION	CD	UC
Unfolded Protein Response, Autophagy and Metabolism					
2q37	<i>ATG16L1</i>	ATG16 autophagy related 16-like 1	Autophagy	+	
5q31	<i>SLC22A5</i>	Solute carrier family 22, member 5	β-Carnitine transporter	+	
5q33	<i>IRGM</i>	Immunity-related GTPase family, M	Autophagy	+	
7p21	<i>AGR2</i>	Anterior gradient 2	Unfolded protein response	+	+
12q12	<i>LRRK2</i>	Leucine-rich repeat kinase 2	Autophagy	+	
13q14	<i>C13orf1</i>	FAMIN/LACC1	Immunometabolic regulator	+	
17q21	<i>ORMDL3</i>	Orosomucoid related member 1-like 3	Unfolded protein response and lipid synthesis	+	+
22q12	<i>XBP1</i>	X-box binding protein 1	Unfolded protein response	+	+
Innate Immunity					
1q23	<i>ITLN1</i>	Intelectin 1	Bacterial binding	+	
16q12	<i>NOD2</i>	Nucleotide-binding oligomerization domain containing 2	Bacterial sensing and autophagy activation	+	
Adaptive Immunity					
1p31	<i>IL23R</i>	Interleukin 23 receptor	T _H 17 cell stimulation	+	+
1q32	<i>IL10</i>	Interleukin 10	Treg-associated cytokine		+
5q33	<i>IL12B</i>	Interleukin 12B	IL-12 p40 chain of IL-12/IL-23	+	+
18p11	<i>PTPN2</i>	Protein tyrosine phosphatase, nonreceptor type 2	T-cell regulation	+	
Inflammation and Healing					
3p21	<i>MST1</i>	Macrophage stimulating 1	Macrophage activation	+	+
5p13	<i>PTGER4</i>	Prostaglandin E receptor 4	PGE ₂ receptor	+	+
6q23	<i>TNFAIP3</i>	Tumor necrosis factor, alpha-induced protein 3 (A20)	Toll-like receptor regulation	+	
6q27	<i>CCR6</i>	Chemokine (C-C motif) receptor 6	Dendritic cell migration	+	
9p24	<i>JAK2</i>	Janus kinase 2	IL-6R and IL-23R signaling	+	+
17q21	<i>STAT3</i>	Signal transducer and activator of transcription 3	IL-6R, IL-23R, and IL-10R signaling	+	+

Abbreviations: CD, Crohn's disease; GTPase, guanosine triphosphatase; IL, interleukin; PGE₂, prostaglandin E₂; Treg, T regulatory cell; UC, ulcerative colitis.

Source: Adapted from A Kaser et al: Ann Rev Immunol 28:573, 2010; Graham DB and Xavier RJ: Nature 578:527, 2020.

healing that control leukocyte recruitment and inflammatory mediator production. Each of these genetic susceptibilities contributes in an incremental manner to IBD risk, variably affects the activities of virtually all subtypes of immune and nonimmune cells within the intestines, and encodes mutations (polymorphisms) that promote or protect from IBD. Some of these loci are associated with specific subtypes of disease such as the association between *NOD2* polymorphisms and fibrostenosing CD or *ATG16L1* and fistulizing disease, especially within the ileum. However, the clinical utility of these genetic risk factors for the diagnosis or determination of prognosis and therapeutic responses remains to be defined.

■ COMMENSAL MICROBIOTA AND IBD

The endogenous commensal microbiota within the intestines plays a central role in the pathogenesis of IBD. Humans are born with sterile guts and acquire their commensal microbiota initially from the mother during egress through the birth canal and subsequently from environmental sources. A stable configuration of up to 1000 species of bacteria that achieves a biomass of $\sim 10^{12}$ colony-forming units per gram of feces is achieved by 3 years of age, which likely persists into adult life, with each individual human possessing a unique combination of species. In addition, the intestines contain other microbial life forms including fungi, archaea, viruses, and protists. The microbiota is thus considered as a critical and sustaining component of the human organism. The establishment and maintenance of the intestinal microbiota composition and function are under the control of host (e.g., immune and epithelial responses), environmental (e.g., diet and antibiotics), and likely genetic (e.g., *NOD2*) factors (Fig. 326-1). In turn, the microbiota, through its structural components and metabolic activity, has major influences on the epithelial and immune function of the host, which, through epigenetic effects, may have durable consequences. During early life when the commensal microbiota is being established, these microbial effects on the host may be particularly important in determining later life risk for IBD. Specific components of the microbiota can promote or protect from disease. The commensal microbiota in patients with both UC and CD is demonstrably different from that of nonafflicted individuals, a state of dysbiosis suggesting the presence of microorganisms that drive disease (e.g., Proteobacteria such as enteroinvasive and adherent *Escherichia coli*) and to which the immune response is directed and/or the loss of microorganisms that hinder inflammation (e.g., Firmicutes such as *Faecalibacterium prausnitzii*). Many of the changes in the commensal microbiota occur as a consequence of the inflammation and are thus potential secondary drivers of disease. In addition, agents that alter the intestinal microbiota such as metronidazole, ciprofloxacin, and elemental diets, may improve CD. CD may also respond to fecal diversion, demonstrating the ability of luminal contents to exacerbate disease.

■ DEFECTIVE IMMUNE REGULATION IN IBD

The mucosal immune system does not normally elicit an inflammatory immune response to luminal contents due to oral (mucosal) tolerance. Administration of soluble antigens orally, rather than subcutaneously or intramuscularly, leads to antigen-specific control of the response and the host's ability to tolerate the antigen. Multiple mechanisms are involved in the induction of oral tolerance and include deletion or anergy (non-responsiveness) of antigen-reactive T cells or induction of CD4+ T cells that suppress gut inflammation (e.g., T regulatory cells expressing the FoxP3 transcription factor) and that secrete anti-inflammatory cytokines such as IL-10, IL-35, and transforming growth factor β (TGF- β). Oral tolerance may be responsible for the lack of immune responsiveness to dietary antigens and the commensal microbiota in the intestinal lumen. In IBD, this suppression of inflammation is altered, leading to uncontrolled inflammation. The mechanisms of this regulated immune suppression are incompletely known.

Gene knockout ($^{-/-}$) or transgenic (Tg) mouse models of IBD, including those that are directed at genes associated with risk for the human disease, have revealed that deleting specific cytokines (e.g., IL-2, IL-10, TGF- β) or their receptors, deleting molecules associated

with T-cell antigen recognition (e.g., T-cell antigen receptors), or interfering with IEC barrier function and the regulation of responses to commensal bacteria (e.g., XBP1, mucus glycoproteins, or nuclear factor- κ B [NF- κ B]) leads to spontaneous colitis or enteritis. In the majority of circumstances, intestinal inflammation in these animal models requires the presence of the commensal microbiota. However, in some cases, activation of certain elements of the intestinal immune system may be exacerbated by the absence of bacteria, resulting in severe colitis and emphasizing the presence of protective properties of the commensal microbiota. Thus, a variety of specific alterations in either the microbiota or host can lead to uncontrolled immune activation and inflammation directed at the intestines in mice. How these relate to human IBD remains to be defined, but they are consistent with inappropriate responses of the genetically susceptible host to the commensal microbiota.

■ THE INFLAMMATORY CASCADE IN IBD

In both UC and CD, inflammation likely emerges from the genetic predisposition of the host in the context of yet-to-be-defined environmental factors. Once initiated in IBD by abnormal innate immune sensing of bacteria by parenchymal cells (e.g., IECs) and hematopoietic cells (e.g., dendritic cells), the immune inflammatory response is perpetuated by T-cell activation when coupled together with inadequate regulatory pathways. A sequential cascade of inflammatory mediators extends the response, making each step a potential target for therapy. Inflammatory cytokines from innate immune cells such as IL-1, IL-6, IL-12, IL-23, and TNF have diverse effects on tissues. They promote fibrogenesis, collagen production, activation of tissue metalloproteinases, and the production of other inflammatory mediators; they also activate the coagulation cascade in local blood vessels (e.g., increased production of von Willebrand factor). These cytokines are normally produced in response to infection but are usually turned off or inhibited by cytokines such as IL-10 and TGF- β at the appropriate time to limit tissue damage. In IBD their activity is not regulated, resulting in an imbalance between the proinflammatory and anti-inflammatory mediators. Some cytokines activate other inflammatory cells (macrophages and B cells), and others act indirectly to recruit other lymphocytes, inflammatory leukocytes, and mononuclear cells from the bloodstream into the gut through interactions between homing receptors on leukocytes (e.g., $\alpha 4\beta 7$ integrin) and addressins on vascular endothelium (e.g., MadCAM1). CD4+ T helper (T_h) cells that promote inflammation are of three major types, all of which may be associated with colitis in animal models and perhaps humans: T_h1 cells (secrete interferon [IFN] γ), T_h2 cells (secrete IL-4, IL-5, IL-13), and T_h17 cells (secrete IL-17, IL-21, IL-22). T_h17 cells may also provide protective functions. Innate immune-like cells (ILCs) that lack T-cell receptors are also present in intestines, polarize to the same functional fates, and may similarly participate in IBD. T_h1 cells induce transmural granulomatous inflammation that resembles CD; T_h2 cells and related natural killer T cells that secrete IL-4, IL-5, and IL-13 induce superficial mucosal inflammation resembling UC in animal models; and T_h17 cells may be responsible for neutrophilic recruitment. However, neutralization of the cytokines produced by these cells, such as IFN- γ or IL-17, has yet to show efficacy in therapeutic trials. Each of these T-cell subsets cross-regulates each other. The T_h1 cytokine pathway is initiated by IL-12, a key cytokine in the pathogenesis of experimental models of mucosal inflammation. IL-4 and IL-23, together with IL-6 and TGF- β , induce T_h2 and T_h17 cells, respectively, and IL-23 inhibits the suppressive function of regulatory T cells. Activated macrophages secrete TNF and IL-6.

These characteristics of the immune response in IBD explain the beneficial therapeutic effects of antibodies to block proinflammatory cytokines or the signaling by their receptors (e.g., anti-TNF, anti-IL-12, anti-IL-23, anti-IL-6, or Janus kinase [JAK] inhibitors) or molecules associated with leukocyte recruitment (e.g., anti- $\alpha 4\beta 7$). They also highlight the potential usefulness of cytokines that inhibit inflammation and promote regulatory T cells or promote intestinal barrier function (e.g., IL-10) in the treatment of IBD. Therapies such as the 5-aminosalicylic acid (5-ASA) compounds and glucocorticoids are also

potent inhibitors of these inflammatory mediators through inhibition of transcription factors such as NF- κ B that regulate their expression.

PATHOLOGY

■ ULCERATIVE COLITIS: MACROSCOPIC FEATURES

UC is a mucosal disease that usually involves the rectum and extends proximally to involve all or part of the colon. About 40–50% of patients have disease limited to the rectum and rectosigmoid, 30–40% have disease extending beyond the sigmoid but not involving the whole colon, and 20% have a pancolitis. Proximal spread occurs in continuity without areas of uninvolved mucosa. When the whole colon is involved, the inflammation extends 2–3 cm into the terminal ileum in 10–20% of patients. The endoscopic changes of *backwash ileitis* are superficial and mild and are of little clinical significance. Although variations in macroscopic activity may suggest skip areas, biopsies from normal-appearing mucosa are usually abnormal. Thus, it is important to obtain multiple biopsies from apparently unininvolved mucosa, whether proximal or distal, during endoscopy. One caveat is that effective medical therapy can change the appearance of the mucosa such that either skip areas or the entire colon can be microscopically normal.

With mild inflammation, the mucosa is erythematous and has a fine granular surface that resembles sandpaper. In more severe disease, the mucosa is hemorrhagic, edematous, and ulcerated (Fig. 326-3). In long-standing disease, inflammatory polyps (pseudopolyps) may be present as a result of epithelial regeneration. The mucosa may appear normal in remission, but in patients with many years of disease, it appears atrophic and featureless, and the entire colon becomes narrowed and shortened. Patients with fulminant disease can develop a toxic colitis or megacolon where the bowel wall becomes thin and the mucosa is severely ulcerated; this may lead to perforation.

■ ULCERATIVE COLITIS: MICROSCOPIC FEATURES

Histologic findings correlate well with the endoscopic appearance and clinical course of UC. The process is limited to the mucosa and superficial submucosa, with deeper layers unaffected except in fulminant disease. In UC, two major histologic features suggest chronicity and help distinguish it from infectious or acute self-limited colitis. First, the crypt architecture of the colon is distorted; crypts may be bifid and reduced in number, often with a gap between the crypt bases and the muscularis mucosae. Second, some patients have basal plasma cells and multiple basal lymphoid aggregates. Mucosal vascular congestion, with edema and focal hemorrhage, and an inflammatory cell infiltrate of neutrophils, lymphocytes, plasma cells, and macrophages may be present. The neutrophils invade the epithelium, usually in the crypts,

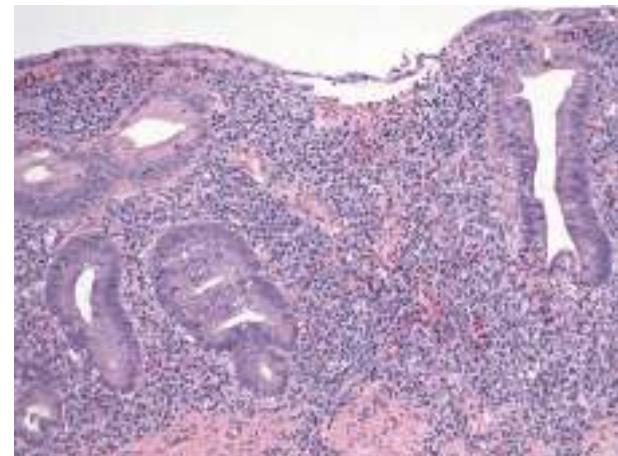


FIGURE 326-4 Medium-power view of colonic mucosa in ulcerative colitis showing diffuse mixed inflammation, basal lymphoplasmacytosis, crypt atrophy and irregularity, and superficial erosion. These features are typical of chronic active ulcerative colitis. (Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

giving rise to cryptitis and, ultimately, to crypt abscesses (Fig. 326-4). Ileal changes in patients with backwash ileitis include villous atrophy and crypt regeneration with increased inflammation, increased neutrophil and mononuclear inflammation in the lamina propria, and patchy cryptitis and crypt abscesses.

■ CROHN'S DISEASE: MACROSCOPIC FEATURES

CD can affect any part of the gastrointestinal (GI) tract from the mouth to the anus. Some 30–40% of patients have small-bowel disease alone, 40–55% have disease involving both the small and large intestines, and 15–25% have colitis alone. In the 75% of patients with small-intestinal disease, the terminal ileum is involved in 90%. Unlike UC, which almost always involves the rectum, the rectum is often spared in CD. CD is often segmental with skip areas throughout the diseased intestine (Fig. 326-5). Perianal disease, manifesting as perirectal fistulas, fissures, abscesses, and anal stenosis, is present in one-third of patients with CD, particularly those with colonic involvement. Rarely, CD may also involve the liver and the pancreas.

Unlike UC, CD is a transmural process. Endoscopically, aphthous or small superficial ulcerations characterize mild disease; in more active disease, stellate ulcerations fuse longitudinally and transversely to



FIGURE 326-3 Ulcerative colitis. Diffuse (nonsegmental) mucosal disease, with broad areas of ulceration. The bowel wall is not thickened, and there is no cobblestoning. (Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)



FIGURE 326-5 Crohn's disease of the colon showing thickening of the wall, with stenosis, linear serpiginous ulcers, and cobblestoning of the mucosa. (Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

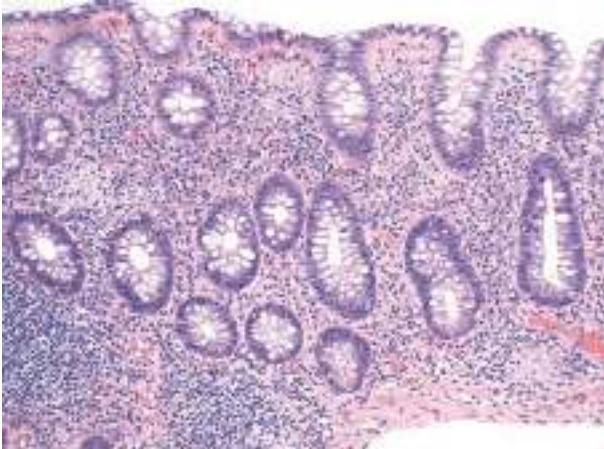


FIGURE 326-6 Medium-power view of Crohn's colitis showing mixed acute and chronic inflammation, crypt atrophy, and multiple small epithelioid granulomas in the mucosa. (Courtesy of Dr. R Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

demarcate islands of mucosa that frequently are histologically normal. This “cobblestone” appearance is characteristic of CD, both endoscopically and by barium radiography. As in UC, pseudopolyps can form in CD.

Active CD is characterized by focal inflammation and formation of fistula tracts, which resolve by fibrosis and stricturing of the bowel. The bowel wall thickens and becomes narrowed and fibrotic, leading to chronic, recurrent bowel obstructions. Projections of thickened mesentery known as “creeping fat” encase the bowel, and serosal and mesenteric inflammation promotes adhesions and fistula formation.

■ CROHN'S DISEASE: MICROSCOPIC FEATURES

The earliest lesions are aphthoid ulcerations and focal crypt abscesses with loose aggregations of macrophages, which form noncaseating granulomas in all layers of the bowel wall (Fig. 326-6). Granulomas are a characteristic feature of CD and are less commonly found on mucosal biopsies than on surgical resection specimens. Other histologic features of CD include submucosal or subserosal lymphoid aggregates, particularly away from areas of ulceration, gross and microscopic skip areas, and transmural inflammation that is accompanied by fissures that penetrate deeply into the bowel wall and sometimes form fistulous tracts or local abscesses.

CLINICAL PRESENTATION

■ ULCERATIVE COLITIS

Signs and Symptoms The major symptoms of UC are diarrhea, rectal bleeding, tenesmus, passage of mucus, and crampy abdominal pain. The severity of symptoms correlates with the extent of disease. Although UC can present acutely, symptoms usually have been present for weeks to months.

Patients with proctitis usually pass fresh blood or blood-stained mucus, either mixed with stool or streaked onto the surface of a normal or hard stool. They also have tenesmus, or urgency with a feeling of incomplete evacuation, but rarely have abdominal pain. With proctitis or proctosigmoiditis, proximal transit slows, which may account for the constipation commonly seen in patients with distal disease.

When the disease extends beyond the rectum, blood is usually mixed with stool or grossly bloody diarrhea may be noted. Colonic motility is altered by inflammation with rapid transit through the inflamed intestine. When the disease is severe, patients pass a liquid stool containing blood, pus, and fecal matter. Diarrhea is often nocturnal and/or postprandial. Although severe pain is not a prominent symptom, some patients with active disease may experience lower abdominal discomfort or mild central abdominal cramping. Severe

TABLE 326-4 Montreal Classification of Extent and Severity of Ulcerative Colitis (UC)

EXTENT	ANATOMY
E1: Ulcerative proctitis	Involvement limited to the rectum
E2: Left-sided UC (distal UC)	Involvement limited to the colorectum distal to the splenic flexure
E3: Extensive UC (pancolitis)	Involvement extends proximal to the splenic flexure
SEVERITY	DEFINITION
S0: Clinical remission	Absence of symptoms
S1: Mild disease activity	≤4 stools/d (with or without blood), absence of systemic illness, normal inflammatory markers (ESR)
S2: Moderate disease activity	≥4 stools/d but minimal signs of systemic toxicity
S3: Severe disease activity	≥6 bloody stools/d, pulse ≥90 beats/min, temperature ≥37.5°C, hemoglobin <10.5 g/100 mL, and ESR ≥30 mm/h

Abbreviation: ESR, erythrocyte sedimentation rate.

Source: C Gasche et al: A simple classification of Crohn's disease: Report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 6:8, 2000; and J Satsangi et al: The Montreal classification of inflammatory bowel disease: Controversies, consensus, and implications. Gut 55:749, 2006.

cramping and abdominal pain can occur with severe attacks of the disease. Other symptoms in moderate to severe disease include anorexia, nausea, vomiting, fever, and weight loss.

Physical signs of proctitis include a tender anal canal and blood on rectal examination. With more extensive disease, patients have tenderness to palpation directly over the colon. Patients with a toxic colitis have severe pain and bleeding, and those with megacolon have hepatic tympany. Both may have signs of peritonitis if a perforation has occurred. The classification of disease activity is shown in Table 326-4.

Laboratory, Endoscopic, and Radiographic Features Active disease can be associated with a rise in acute-phase reactants (C-reactive protein [CRP]), platelet count, and erythrocyte sedimentation rate (ESR) and a decrease in hemoglobin. Fecal lactoferrin, a glycoprotein present in activated neutrophils, is a highly sensitive and specific marker for detecting intestinal inflammation. Fecal calprotectin is present in neutrophils and monocytes, and levels correlate well with histologic inflammation, predict relapses, and detect pouchitis. Both fecal lactoferrin and calprotectin are becoming an integral part of IBD management and are used frequently to rule out active inflammation versus symptoms of irritable bowel or bacterial overgrowth. In severely ill patients, the serum albumin level will fall rather quickly. Leukocytosis may be present but is not a specific indicator of disease activity. Proctitis or proctosigmoiditis rarely causes a rise in CRP. Diagnosis relies on the patient's history, clinical symptoms, negative stool and/or tissue examination for bacteria, *C. difficile* toxin, ova and parasites, and viruses depending on epidemiologic considerations and clinical presentation; sigmoidoscopic appearance (see Fig. 322-4A); and histology of rectal or colonic biopsy specimens.

Sigmoidoscopy is used to assess disease activity and is usually performed before treatment. If the patient is not having an acute flare, colonoscopy is used to assess disease extent and activity (Fig. 326-7). Endoscopically mild disease is characterized by erythema, decreased vascular pattern, and mild friability. Moderate disease is characterized by marked erythema, absent vascular pattern, friability, and erosions, and severe disease is characterized by spontaneous bleeding and ulcerations. Histologic features change more slowly than clinical features but can also be used to grade disease activity.

Complications Only 15% of patients with UC present initially with severe disease. Massive hemorrhage occurs in 1% of patients, and treatment for the disease usually stops the bleeding. *Toxic megacolon* is defined as a transverse or right colon with a diameter of >6 cm, with loss of haustration in patients with severe attacks of UC. It occurs rarely



FIGURE 326-7 Colonoscopy with acute ulcerative colitis: severe colon inflammation with erythema, friability, and exudates. (Courtesy of Dr. M. Hamilton, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

and can be triggered by electrolyte abnormalities and narcotics. About 50% of acute dilations will resolve with conservative management alone, but urgent colectomy is required for those who do not improve. Perforation is the most dangerous of the local complications, and the physical signs of peritonitis may not be obvious, especially if the patient is receiving glucocorticoids. Although perforation is rare, the mortality rate for perforation complicating a toxic megacolon is ~15%. In addition, patients can develop a toxic colitis and such severe ulcerations that the bowel may perforate without first dilating.

Strictures occur in 5–10% of patients and are always a concern in UC because of the possibility of underlying neoplasia. Although benign strictures can form from the inflammation and fibrosis of UC, strictures that are impassable with the colonoscope should be presumed malignant until proven otherwise. A stricture that prevents passage of the colonoscope is an indication for surgery. UC patients occasionally develop anal fissures, perianal abscesses, or hemorrhoids, but the occurrence of extensive perianal lesions should suggest CD.

CROHN'S DISEASE

Signs and Symptoms Although CD usually presents as acute or chronic bowel inflammation, the inflammatory process evolves toward one of two patterns of disease: a fibrostenotic obstructing pattern or a penetrating fistulous pattern, each with different treatments and prognoses. The site of disease influences the clinical manifestations (Table 326-5).

TABLE 326-5 Vienna and Montreal Classifications of Crohn's Disease

	VIENNA	MONTREAL
Age at diagnosis	A1: <40 years A2: >40 years	A1: <16 years A2: Between 17 and 40 years A3: >40 years
Location	L1: Ileal L2: Colonic L3: Ileocolonic L4: Upper	L1: Ileal L2: Colonic L3: Ileocolonic L4: Isolated upper disease ^a
Behavior	B1: Nonstricturing, nonpenetrating B2: Stricturing B3: Penetrating p: Perianal disease modifier ^b	B1: Nonstricturing, nonpenetrating B2: Stricturing B3: Penetrating

^aL4 is a modifier and can be added to L1–L3 when there is concomitant foregut disease.

^bp is added to B1–B3 when there is concomitant perianal disease.

ILEOCOLITIS Because the most common site of inflammation is the terminal ileum, the usual presentation of ileocolitis is a chronic history of recurrent episodes of right lower quadrant pain and diarrhea. Sometimes the initial presentation mimics acute appendicitis with pronounced right lower quadrant pain, a palpable mass, fever, and leukocytosis. Pain is usually colicky; it precedes and is relieved by defecation. A low-grade fever is usually noted. High-spiking fever suggests intraabdominal abscess formation. Weight loss is common—typically 10–20% of body weight—and develops as a consequence of diarrhea, anorexia, and fear of eating.

An inflammatory mass may be palpated in the right lower quadrant of the abdomen. The mass is composed of inflamed bowel, induration of the mesentery, and enlarged abdominal lymph nodes. The “string sign” on radiographic studies results from a severely narrowed loop of bowel, which makes the lumen resemble a frayed cotton string. It is caused by incomplete filling of the lumen as the result of edema, irritability, and spasms associated with inflammation and ulcerations. The sign may be seen in both nonstenotic and stenotic phases of the disease.

Bowel obstruction may take several forms. In the early stages of disease, bowel wall edema and spasm produce intermittent obstructive manifestations and increasing symptoms of postprandial pain. Over several years, persistent inflammation gradually progresses to fibrostenotic narrowing and stricture. Diarrhea will decrease and be replaced by chronic bowel obstruction. Acute episodes of obstruction occur as well, precipitated by bowel inflammation and spasm or sometimes by impaction of undigested food or medication. These episodes usually resolve with intravenous fluids and gastric decompression.

Severe inflammation of the ileocecal region may lead to localized wall thinning, with microperforation and fistula formation to the adjacent bowel, the skin, or the urinary bladder, or to an abscess cavity in the mesentery. Enterovesical fistulas typically present as dysuria or recurrent bladder infections or, less commonly, as pneumaturia or fecaluria. Enterocutaneous fistulas follow tissue planes of least resistance, usually draining through abdominal surgical scars. Enterovaginal fistulas are rare and present as dyspareunia or as a feculent or foul-smelling, often painful vaginal discharge. They are unlikely to develop without a prior hysterectomy.

JEJUNOILEITIS Extensive inflammatory disease is associated with a loss of digestive and absorptive surface, resulting in malabsorption and steatorrhea. Nutritional deficiencies can also result from poor intake and enteric losses of protein and other nutrients. Intestinal malabsorption can cause anemia, hypoalbuminemia, hypocalcemia, hypomagnesemia, coagulopathy, and hyperoxaluria with nephrolithiasis in patients with an intact colon. Many patients need to take intravenous iron since oral iron is poorly tolerated and often ineffective. Vertebral fractures are caused by a combination of vitamin D deficiency, hypocalcemia, and prolonged glucocorticoid use. Pellagra from niacin deficiency can occur in extensive small-bowel disease, and malabsorption of vitamin B₁₂ can lead to megaloblastic anemia and neurologic symptoms. Other important nutrients to measure and replete if low are folate and vitamins A, E, and K. Levels of minerals such as zinc, selenium, copper, and magnesium are often low in patients with extensive small-bowel inflammation or resections, and these should be repleted as well. Most patients should take daily multivitamin, calcium, and vitamin D supplements.

Diarrhea is characteristic of active disease; its causes include (1) bacterial overgrowth in obstructive stasis or fistulization, (2) bile acid malabsorption due to a diseased or resected terminal ileum, (3) intestinal inflammation with decreased water absorption and increased secretion of electrolytes and (4) enteroenteric fistula(e).

COLITIS AND PERIANAL DISEASE Patients with colitis present with low-grade fevers, malaise, diarrhea, crampy abdominal pain, and sometimes hematochezia. Gross bleeding is not as common as in UC and appears in about one-half of patients with exclusively colonic disease. Only 1–2% exhibit massive bleeding. Pain is caused by passage of fecal material through narrowed and inflamed segments of the large bowel. Decreased rectal compliance is another cause for diarrhea in Crohn's colitis patients.

Stricture can occur in the colon in 4–16% of patients and produce symptoms of bowel obstruction. If the endoscopist is unable to traverse a stricture in Crohn's colitis, surgical resection should be considered, especially if the patient has symptoms of chronic obstruction. Colonic disease may fistulize into the stomach or duodenum, causing feculent vomiting, or to the proximal or mid-small bowel, causing malabsorption by "short circuiting" the absorptive surface and bacterial overgrowth. Ten percent of women with Crohn's colitis will develop a rectovaginal fistula.

Perianal disease affects about one-third of patients with Crohn's colitis and is manifested by incontinence, large hemorrhoidal tags, anal strictures, anorectal fistulas, and perirectal abscesses. Not all patients with perianal fistula will have endoscopic evidence of colonic inflammation.

GASTRODUODENAL DISEASE Symptoms and signs of upper GI tract disease include nausea, vomiting, and epigastric pain. Patients usually have a *Helicobacter pylori*-negative gastritis. The second portion of the duodenum is more commonly involved than the bulb. Fistulas involving the stomach or duodenum arise from the small or large bowel and do not necessarily signify the presence of upper GI tract involvement. Patients with advanced gastroduodenal CD may develop a chronic gastric outlet obstruction. About 30% of children diagnosed with CD have esophagogastrroduodenal involvement. The classification of disease activity is shown in Table 326-5.

Laboratory, Endoscopic, and Radiographic Features Laboratory abnormalities include elevated ESR and CRP. In more severe disease, findings include hypoalbuminemia, anemia, and leukocytosis. Fecal calprotectin and lactoferrin levels have been used to distinguish IBD from irritable bowel syndrome (IBS), to assess whether CD is active, and to detect postoperative recurrence of CD. Fecal calprotectin is a more sensitive marker of ileocolonic or colonic inflammation rather than isolated ileal inflammation.

Endoscopic features of CD include rectal sparing, aphthous ulcerations, fistulas, and skip lesions. Colonoscopy allows examination and biopsy of mass lesions or strictures and biopsy of the terminal ileum. Upper endoscopy is useful in diagnosing gastroduodenal involvement in patients with upper tract symptoms. Ileal or colonic strictures may be dilated with balloons introduced through the colonoscope. Strictures ≤ 4 cm in length and those at anastomotic sites respond better to endoscopic dilation. The perforation rate is as high as 10%. Most endoscopists dilate only fibrotic strictures and not those associated with active inflammation. Wireless capsule endoscopy (WCE) allows direct visualization of the entire small-bowel mucosa (Fig. 326-8). The diagnostic yield of detecting lesions suggestive of active CD is higher with WCE than CT or magnetic resonance (MR) enterography. WCE should be used in the setting of a small-bowel stricture. Capsule retention occurs in <1% of patients with suspected CD, but retention rates of 4–6% are seen in patients with established CD. It is helpful to give the patient with CD a patency capsule, which is made of barium and starts to dissolve 30 h after ingestion. An abdominal x-ray can be taken at around 30 h after ingestion to see if the capsule is still present in the small bowel, which would indicate a stricture.

In CD, early radiographic findings in the small bowel include thickened folds and aphthous ulcerations. "Cobblestoning" from longitudinal and transverse ulcerations most frequently involves the small bowel. In more advanced disease, strictures, fistulas, inflammatory masses, and abscesses may be detected. The earliest macroscopic findings of colonic CD are aphthous ulcers. These small ulcers are often multiple and separated by normal intervening mucosa. As the disease progresses, aphthous ulcers become enlarged, deeper, and occasionally connected to one another, forming longitudinal stellate, serpiginous, and linear ulcers (see Fig. 322-4B).

The transmural inflammation of CD leads to decreased luminal diameter and limited distensibility. As ulcers progress deeper, they can lead to fistula formation. The segmental nature of CD results in wide gaps of normal or dilated bowel between involved segments.

CT enterography and MR enterography have been shown to be equally accurate in the identification of active small-bowel



FIGURE 326-8 Wireless capsule endoscopy image in a patient with Crohn's disease of the ileum shows ulcerations and narrowing of the intestinal lumen. (Courtesy of Dr. S. Reddy, Gastroenterology Division, Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

inflammation. MRI is thought to offer superior soft tissue contrast and has the added advantage of avoiding radiation exposure changes (Figs. 326-9 and 326-10). The lack of ionizing radiation is particularly appealing in younger patients and when monitoring response to therapy where serial images will be obtained. Pelvic MRI is superior to pelvic CT for demonstrating pelvic lesions such as ischiorectal abscesses and perianal fistulas (Fig. 326-11). An underutilized resource for assessing small-bowel CD is small-bowel ultrasound (SBUS). SBUS is at least as sensitive as MR enterography and CT enterography for detecting small-bowel CD, with a sensitivity of 94%, specificity of 97%, positive predictive value of 97%, and negative predictive value of 94%. Use of oral contrast medium can increase the sensitivity and specificity to detect small-bowel lesions to 100%. SBUS is best suited for distal small-bowel assessment, as the sensitivity of detecting lesions within the duodenum and proximal jejunum may be lower due to anatomic position. The limitations of SBUS include availability and operator dependence.

Complications Because CD is a transmural process, serosal adhesions develop that provide direct pathways for fistula formation and reduce the incidence of free perforation. Perforation occurs in 1–2% of patients, usually in the ileum but occasionally in the jejunum or as a complication of toxic megacolon. The peritonitis of free perforation, especially colonic, may be fatal. Intraabdominal and pelvic abscesses occur in 10–30% of patients with CD at some time in the course of their illness. CT-guided percutaneous drainage of the abscess is standard therapy. Despite adequate drainage, most patients need resection of the offending bowel segment. Percutaneous drainage has an especially high failure rate in abdominal wall abscesses. Systemic glucocorticoid therapy increases the risk of intraabdominal and pelvic abscesses in CD patients who have never had an operation. Other complications include intestinal obstruction in 40%, massive hemorrhage, malabsorption, and severe perianal disease.

Serologic Markers Patients with UC and CD show a wide variation in the way they present and progress over time. Some patients present with mild disease activity and do well with generally safe and mild medications, but many others exhibit more severe disease and can develop serious complications that will require surgery. Current and developing biologic therapies can help halt progression of disease



FIGURE 326-9 A coronal magnetic resonance image was obtained using a half Fourier single-shot T2-weighted acquisition with fat saturation in a 27-year-old pregnant (23 weeks' gestation) woman. The patient had Crohn's disease and was maintained on mercaptopurine and prednisone. She presented with abdominal pain, distension, vomiting, and small-bowel obstruction. The image reveals a 7- to 10-cm long stricture at the terminal ileum (white arrows) causing obstruction and significant dilatation of the proximal small bowel (white asterisk). A fetus is seen in the uterus (dashed white arrows). (Courtesy of Drs. J. F. B. Chick and P. B. Shyn, *Abdominal Imaging and Intervention*, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.)

and give patients with moderate to severe UC and CD a better quality of life. There are potential risks of biologic therapies such as infection and malignancy, and it would be optimal to determine by genetic or serologic markers at the time of diagnosis which patients will require more aggressive medical therapy.

For success in diagnosing IBD and in differentiating between CD and UC, the efficacy of these serologic tests depends on the prevalence of IBD in a specific population. Increased titers of anti-*Saccharomyces cerevisiae* antibody (ASCA) have been associated with CD, whereas increased levels of perinuclear antineutrophil cytoplasmic antibody (pANCA) are more commonly seen in patients with UC. However, when evaluated in a meta-analysis of 60 studies, the sensitivity and specificity of an ASCA-positive/pANCA-negative pattern for identification of CD were 55 and 93%, respectively. In addition to ASCA, multiple other antibodies to bacterial proteins (Omp-C and I2), flagellin (CBir1), and bacterial carbohydrates have been studied and associated with CD. These serologic markers tend to have low sensitivity and specificity and may be elevated due to other autoimmune diseases, infections, and inflammation including those outside of the GI tract. The Prometheus IBD SGI Diagnostic blood test measures a panel of serologic (S), genetic (G), and inflammatory (I) biomarkers, but the test is costly, and reliable results are based on the pretest probability of the patient having IBD. PROSPECT is a validated web-based tool to display individual CD outcomes and considers multiple variables including disease location (large or small bowel, perianal), serologies (ASCA, CBir1, ANCA), and genetics (*NOD2* frameshift mutation).

Clinical factors described at diagnosis are more helpful than serologies at predicting the natural history of IBD. Except in special circumstances (such as before consideration of an ileal pouch-anal anastomosis [IPAA] in a patient with indeterminate colitis), serologic markers have only minimal clinical utility.

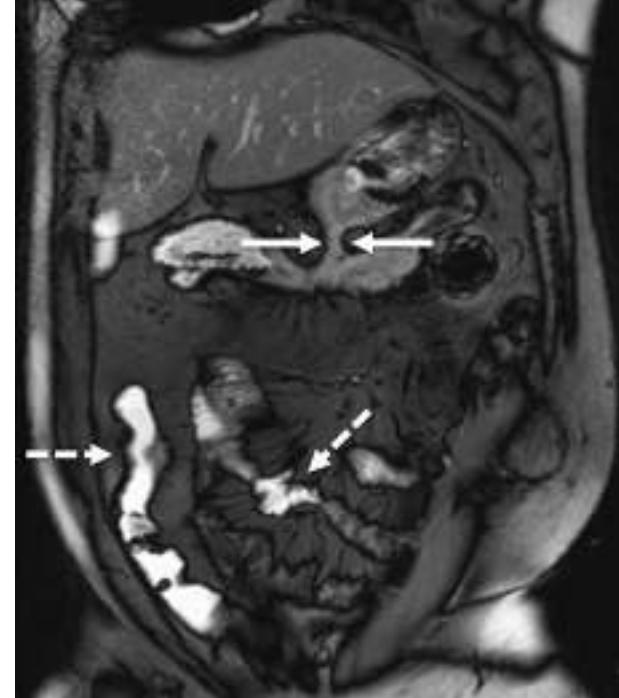


FIGURE 326-10 A coronal balanced, steady-state, free precession, T2-weighted image with fat saturation was obtained in a 32-year-old man with Crohn's disease and prior episodes of bowel obstruction, fistulas, and abscesses. He was being treated with mercaptopurine and presented with abdominal distention and diarrhea. The image demonstrates a new gastrocolic fistula (solid white arrows). Multifocal involvement of the small bowel and terminal ileum is also present (dashed white arrows). (Courtesy of Drs. J. F. B. Chick and P. B. Shyn, *Abdominal Imaging and Intervention*, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.)

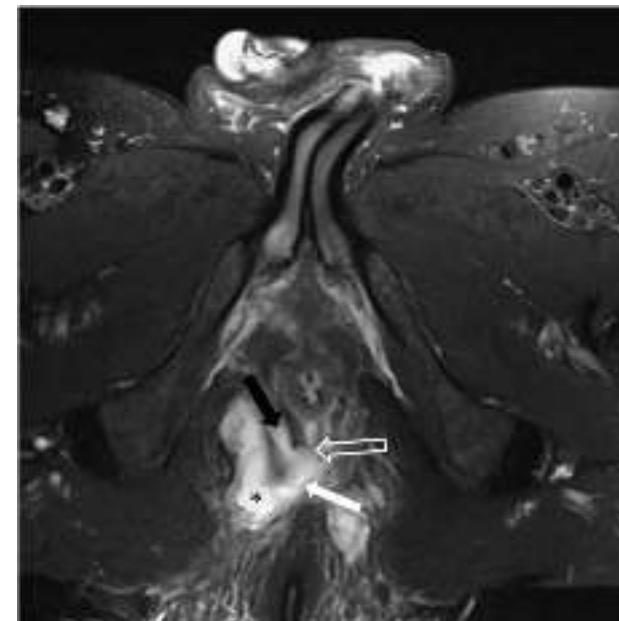


FIGURE 326-11 Axial T2-weighted fat-saturated image obtained in a 39-year-old male with Crohn's disease shows a defect in the internal sphincter at the 6 o'clock position of the mid anal canal (open white arrow) communicating with a 1.1-cm intersphincteric collection (black arrow). Wide defect in the external sphincter at the 7 o'clock position (solid white arrow) leads to a moderate-sized perianal abscess in the ischioanal fossa (asterisk). (Courtesy of Drs. J. S. Quon and P. B. Shyn, *Abdominal Imaging and Intervention*, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; with permission.)

DIFFERENTIAL DIAGNOSIS OF UC AND CD

Once a diagnosis of IBD is made, distinguishing between UC and CD is impossible initially in up to 15% of cases. These are termed *indeterminate colitis*. Fortunately, in most cases, the true nature of the underlying colitis becomes evident later in the course of the patient's disease. Approximately 5% (range 1–20%) of colon resection specimens are difficult to classify as either UC or CD because they exhibit overlapping histologic features.

■ INFECTIOUS DISEASES

Infections of the small intestines and colon can mimic CD or UC. They may be bacterial, fungal, viral, or protozoal in origin (Table 326-6). *Campylobacter* colitis can mimic the endoscopic appearance of severe UC and can cause a relapse of established UC. *Salmonella* can cause watery or bloody diarrhea, nausea, and vomiting. Shigellosis causes watery diarrhea, abdominal pain, and fever followed by rectal tenesmus and by the passage of blood and mucus per rectum. All three are usually self-limited, but 1% of patients infected with *Salmonella* become asymptomatic carriers. *Yersinia enterocolitica* infection occurs mainly in the terminal ileum and causes mucosal ulceration, neutrophil invasion, and thickening of the ileal wall. Other bacterial infections that may mimic IBD include *C. difficile*, which presents with watery diarrhea, tenesmus, nausea, and vomiting; and *E. coli*, three categories of which can cause colitis. These are enterohemorrhagic, enteroinvasive, and enteroadherent *E. coli*, all of which can cause bloody diarrhea and abdominal tenderness. Gonorrhea, *Chlamydia*, and syphilis can also cause proctitis.

GI involvement with mycobacterial infection occurs primarily in the immunosuppressed patient but may occur in patients with normal immunity. Distal ileal and cecal involvement predominates, and patients present with symptoms of small-bowel obstruction and

a tender abdominal mass. The diagnosis is made most directly by colonoscopy with biopsy and culture. Although most of the patients with viral colitis are immunosuppressed, cytomegalovirus (CMV) and herpes simplex proctitis may occur in immunocompetent individuals. CMV occurs most commonly in the esophagus, colon, and rectum but may also involve the small intestine. Symptoms include abdominal pain, bloody diarrhea, fever, and weight loss. With severe disease, necrosis and perforation can occur. Diagnosis is made by identification of characteristic intranuclear inclusions in mucosal cells on biopsy. Herpes simplex infection of the GI tract is limited to the oropharynx, anorectum, and perianal areas. Symptoms include anorectal pain, tenesmus, constipation, inguinal adenopathy, difficulty with urinary voiding, and sacral paresthesias. Diagnosis is made by rectal biopsy with identification of characteristic cellular inclusions and viral culture. HIV itself can cause diarrhea, nausea, vomiting, and anorexia. Small-intestinal biopsies show partial villous atrophy; small-bowel bacterial overgrowth and fat malabsorption may also be noted.

Protozoan parasites include *Isospora belli*, which can cause a self-limited infection in healthy hosts but causes a chronic profuse, watery diarrhea and weight loss in AIDS patients. *Entamoeba histolytica* or related species infect ~10% of the world's population; symptoms include abdominal pain, tenesmus, frequent loose stools containing blood and mucus, and abdominal tenderness. Colonoscopy reveals focal punctate ulcers with normal intervening mucosa; diagnosis is made by biopsy or serum amebic antibodies. Fulminant amebic colitis is rare but has a mortality rate of >50%.

Other parasitic infections that may mimic IBD include hookworm (*Necator americanus*), whipworm (*Trichuris trichiura*), and *Strongyloides stercoralis*. In severely immunocompromised patients, *Candida* or *Aspergillus* can be identified in the submucosa. Disseminated histoplasmosis can involve the ileocecal area.

■ NONINFECTIOUS DISEASES

Diverticulitis can be confused with CD clinically and radiographically. Both diseases cause fever, abdominal pain, tender abdominal mass, leukocytosis, elevated ESR, partial obstruction, and fistulas. Perianal disease or ileitis on small-bowel series favors the diagnosis of CD. Significant endoscopic mucosal abnormalities are more likely in CD than in diverticulitis. Endoscopic or clinical recurrence following segmental resection favors CD. Diverticular-associated colitis is similar to CD, but mucosal abnormalities are limited to the sigmoid and descending colon.

Ischemic colitis is commonly confused with IBD. The ischemic process can be chronic and diffuse, as in UC, or segmental, as in CD. Colonic inflammation due to ischemia may resolve quickly or may persist and result in transmural scarring and stricture formation. Ischemic bowel disease should be considered in the elderly following abdominal aortic aneurysm repair or when a patient has a hypercoagulable state or a severe cardiac or peripheral vascular disorder. Patients usually present with sudden onset of left lower quadrant pain, urgency to defecate, and the passage of bright red blood per rectum. Endoscopic examination often demonstrates a normal-appearing rectum and a sharp transition to an area of inflammation in the descending colon and splenic flexure.

The effects of radiotherapy on the GI tract can be difficult to distinguish from IBD. Acute symptoms can occur within 1–2 weeks of starting radiotherapy. When the rectum and sigmoid are irradiated, patients develop bloody, mucoid diarrhea and tenesmus, as in distal UC. With small-bowel involvement, diarrhea is common. Late symptoms include malabsorption and weight loss. Strictureting with obstruction and bacterial overgrowth may occur. Fistulas can penetrate the bladder, vagina, or abdominal wall. Flexible sigmoidoscopy reveals mucosal granularity, friability, numerous telangiectasias, and occasionally discrete ulcerations. Biopsy can be diagnostic.

Solitary rectal ulcer syndrome is uncommon and can be confused with IBD. It occurs in persons of all ages and may be caused by impaired evacuation and failure of relaxation of the puborectalis muscle. Single or multiple ulcerations may arise from anal sphincter overactivity, higher intrarectal pressures during defecation, and digital

TABLE 326-6 Diseases That Mimic IBD

Infectious Etiologies

Bacterial	Mycobacterial	Viral
<i>Salmonella</i>	Tuberculosis	Cytomegalovirus
<i>Shigella</i>	<i>Mycobacterium avium</i>	Herpes simplex
Toxigenic	Parasitic	HIV
<i>Escherichia coli</i>	Amebiasis	Fungal
<i>Campylobacter</i>	<i>Isospora</i>	Histoplasmosis
<i>Yersinia</i>	<i>Trichuris trichiura</i>	<i>Candida</i>
<i>Clostridium difficile</i>	Hookworm	<i>Aspergillus</i>
Gonorrhea	<i>Strongyloides</i>	
<i>Chlamydia trachomatis</i>		

Noninfectious Etiologies

Inflammatory	Neoplastic	Drugs and Chemicals
Appendicitis	Lymphoma	NSAIDs
Diverticulitis	Metastatic	Phosphosoda
Diversion colitis	Carcinoma	Cathartic colon
Collagenous/lymphocytic colitis	Carcinoma of the ileum	Gold
Ischemic colitis	Carcinoid	Oral contraceptives
Radiation colitis/enteritis	Familial polyposis	Cocaine
Solitary rectal ulcer syndrome		Immune checkpoint inhibitor colitis
Eosinophilic gastroenteritis		Mycophenolate mofetil
Neutropenic colitis		
Behçet's syndrome		
Graft-versus-host disease		

Abbreviations: IBD, inflammatory bowel disease; NSAIDs, nonsteroidal anti-inflammatory drugs.

removal of stool. Patients complain of constipation with straining and pass blood and mucus per rectum. Other symptoms include abdominal pain, diarrhea, tenesmus, and perineal pain. Ulceration, which may be as large as 5 cm in diameter, is usually observed anteriorly or anterolaterally 3–15 cm from the anal verge. Biopsies can be diagnostic.

Several types of colitis are associated with nonsteroidal anti-inflammatory drugs (NSAIDs), including de novo colitis, reactivation of IBD, and proctitis caused by use of suppositories. Most patients with NSAID-related colitis present with diarrhea and abdominal pain, and complications include stricture, bleeding, obstruction, perforation, and fistulization. Withdrawal of these agents is crucial, and in cases of reactivated IBD, standard therapies are indicated.

Colitis secondary to immune checkpoint inhibitors (ICIs), termed ICI-related colitis, has emerged as these agents have found use in a wide variety of cancers. Immune checkpoint proteins such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death protein 1 (PD-1) are receptors expressed on the surface of effector T cells that interact with their ligands CD80/CD86 (CTLA-4) and programmed death-ligand 1 (PD-1) on antigen-presenting cells and normally function as inhibitors of immune responses. ICIs block these inhibitory pathways and promote the activation and proliferation of the native adaptive T-cell response against malignant cells as their mechanism of antitumor activity. While very effective at enhancing antitumor T-cell activity, ICIs also activate global T-cell responses that induce several autoimmune-related adverse events. Although immune-related adverse events of ICIs occur in multiple organ systems, the GI tract is affected in 21–44% of patients. The most common clinical presentation is self-limited diarrhea that can be associated with frank colitis and can lead to significant morbidity and mortality if not managed appropriately. Treatment is generally based on symptom severity. Moderate to severe symptoms usually require glucocorticoids, whereas biologics such as anti-TNF agents and integrin inhibitors are used in steroid-refractory cases.

■ THE ATYPICAL COLITIDES

Two atypical colitides—collagenous colitis and lymphocytic colitis—have completely normal endoscopic appearances. Collagenous colitis has two main histologic components: increased subepithelial collagen deposition and colitis with increased intraepithelial lymphocytes. The female-to-male ratio is 9:1, and most patients present in the sixth or seventh decade of life. The main symptom is chronic watery diarrhea. Risk factors include smoking; use of NSAIDs, proton pump inhibitors, or beta blockers; and a history of autoimmune disease.

Lymphocytic colitis has features similar to collagenous colitis, including age at onset and clinical presentation, but it has almost equal incidence in men and women and no subepithelial collagen deposition on pathologic section. However, intraepithelial lymphocytes are increased. Use of sertraline (but not beta blockers) is an additional risk factor. The frequency of celiac disease is increased in lymphocytic colitis and ranges from 9 to 27%. Celiac disease should be excluded in all patients with lymphocytic colitis, particularly if diarrhea does not respond to conventional therapy. Treatments for both microscopic colitides vary depending on symptom severity and include, antidiarrheals (e.g., loperamide and diphenoxylate), bismuth, aminosalicylates, budesonide, systemic glucocorticoids, and biologics for refractory disease.

Diversion colitis is an inflammatory process that arises in segments of the large intestine that are not continuous with the fecal stream. It usually occurs in patients with ileostomy or colostomy when a mucus fistula or a Hartmann's pouch has been created. Clinically, patients have mucus or bloody discharge from the rectum. Erythema, granularity, friability, and, in more severe cases, ulceration can be seen on endoscopy. Histopathology shows areas of active inflammation with foci of cryptitis and crypt abscesses. Crypt architecture is normal, which differentiates it from UC but not necessarily CD. Short-chain fatty acid enemas may help in diversion colitis, but the definitive therapy is surgical reanastomosis.

EXTRAINTESTINAL MANIFESTATIONS

Up to one-third of IBD patients have at least one extraintestinal disease manifestation. Please see Table 326-7 for a summary of IBD EIMs.

■ DERMATOLOGIC

Erythema nodosum (EN) occurs in up to 15% of CD patients and 10% of UC patients. Attacks usually correlate with bowel activity; skin lesions develop after the onset of bowel symptoms, and patients frequently have concomitant active peripheral arthritis. The lesions of EN are hot, red, tender nodules measuring 1–5 cm in diameter and are found on the anterior surface of the lower legs, ankles, calves, thighs, and arms. Therapy is directed toward the underlying bowel disease.

Pyoderma gangrenosum (PG) is seen in 1–12% of UC patients and less commonly in Crohn's colitis. Although it usually presents after the diagnosis of IBD, PG may occur years before the onset of bowel symptoms, run a course independent of the bowel disease, respond poorly to colectomy, and even develop years after proctocolectomy. It is usually associated with severe disease. Lesions are commonly found on the dorsal surface of the feet and legs but may occur on the arms, chest, stoma, and even the face. PG usually begins as a pustule and then spreads concentrically to rapidly undermine healthy skin. Lesions then ulcerate, with violaceous edges surrounded by a margin of erythema. Centrally, they contain necrotic tissue with blood and exudates. Lesions may be single or multiple and grow as large as 30 cm. They are sometimes very difficult to treat and often require IV antibiotics, IV glucocorticoids, dapsone, azathioprine, thalidomide, IV cyclosporine (CSA), infliximab, or adalimumab.

Other dermatologic manifestations include pyoderma vegetans, which occurs in intertriginous areas; pyostomatitis vegetans, which involves the mucous membranes; Sweet syndrome, a neutrophilic dermatosis; and metastatic CD, a rare disorder defined by cutaneous granuloma formation. Psoriasis affects 5–10% of patients with IBD and is unrelated to bowel activity, consistent with the potential shared immunogenetic basis of these diseases. Perianal skin tags are found in 75–80% of patients with CD, especially those with colon involvement. Oral mucosal lesions, seen often in CD and rarely in UC, include aphthous stomatitis and “cobblestone” lesions of the buccal mucosa.

■ RHEUMATOLOGIC

Peripheral arthritis develops in 15–20% of IBD patients, is more common in CD, and worsens with exacerbations of bowel activity. It is asymmetric, polyarticular, and migratory and most often affects large joints of the upper and lower extremities. Treatment is directed at reducing bowel inflammation. In severe UC, colectomy frequently cures the arthritis.

Ankylosing spondylitis (AS) occurs in ~10% of IBD patients and is more common in CD than UC. About two-thirds of IBD patients with AS express the HLA-B27 antigen. The AS activity is not related to bowel activity and does not remit with glucocorticoids or colectomy. It most often affects the spine and pelvis, producing symptoms of diffuse low-back pain, buttock pain, and morning stiffness. The course is continuous and progressive, leading to permanent skeletal damage and deformity. Anti-TNF therapy reduces spinal inflammation and improves functional status and quality of life.

Sacroiliitis is symmetric, occurs equally in UC and CD, is often asymptomatic, does not correlate with bowel activity, and does not always progress to AS. Other rheumatic manifestations include hypertrophic osteoarthropathy, pelvic/femoral osteomyelitis, and relapsing polychondritis.

■ OCULAR

The incidence of ocular complications in IBD patients is 1–10%. The most common are conjunctivitis, anterior uveitis/iritis, and episcleritis. Uveitis is associated with both UC and Crohn's colitis, may be found during periods of remission, and may develop in patients following bowel resection. Symptoms include ocular pain, photophobia, blurred vision, and headache. Prompt intervention, sometimes with systemic glucocorticoids, is required to prevent scarring and visual impairment. Episcleritis is a benign disorder that presents with symptoms of mild

TABLE 326-7 Extraintestinal Manifestations

CATEGORY	CLINICAL COURSE	TREATMENT
Rheumatologic disorders (5–20%)		
Peripheral arthritis	Asymmetric, migratory Parallels bowel activity	Reduce bowel inflammation
Sacroiliitis	Symmetric: spine and hip joints Independent of bowel activity	Steroids, injections, methotrexate, anti-TNF
Ankylosing spondylitis	Gradual fusion of spine Independent of bowel activity Two-thirds have HLA-B27 antigen	Physical therapy, steroids, injections, methotrexate, anti-TNF, IL-17 inhibitors, tofacitinib
Metabolic bone disorders (up to 40% of patients)		
Osteoporosis	Risk increased by glucocorticoids, cyclosporine, methotrexate, total parenteral nutrition, malabsorption, and inflammation Fracture rates highest in the elderly (age >60)	Screening with DEXA scan, check vitamin D levels, treat if osteoporosis or osteopenia on long-term corticosteroids
Osteonecrosis	Death of osteocytes and adipocytes and eventual bone collapse; affects hips more than knees or shoulders; risk factor is steroid use	Pain control, injections, joint replacement
Dermatologic disorders (10–20%)		
Erythema nodosum	Hot, red, tender, nodules/extremities Parallels bowel activity	Reduce bowel inflammation
Pyoderma gangrenosum	Ulcerating, necrotic lesions on extremities, trunk, face, stoma Independent of bowel activity	Antibiotics, steroids, cyclosporine, infliximab, dapsone, azathioprine, intralesional steroids; not debridement or colectomy
Psoriasis	Unrelated to bowel activity	Topical steroids, light therapy, methotrexate, infliximab, adalimumab, ustekinumab
Pyoderma vegetans	Intertriginous areas Parallels bowel activity	Evanescence; resolves without progression
Pyostomatitis vegetans	Mucous membranes Parallels bowel activity	Evanescence; resolves without progression
Metastatic Crohn's disease (CD)	CD of the skin Parallels bowel activity	Reduce bowel inflammation
Sweet syndrome	Neutrophilic dermatosis Parallels bowel activity	Reduce bowel inflammation
Aphthous stomatitis	Oral ulcerations Parallels bowel activity	Reduce bowel inflammation/topical therapy
Ocular disorders (1–11%)		
Uveitis	Ocular pain, photophobia, blurred vision, headache Independent of bowel activity	Topical or systemic steroids
Episcleritis	Mild ocular burning Parallels bowel activity	Topical corticosteroids
Hepatobiliary disorders (10–35%)		
Fatty liver	Secondary to chronic illness, malnutrition, steroid therapy	Improve nutrition, reduce steroids
Cholelithiasis	Patients with ileitis or ileal resection Malabsorption of bile acids, depletion of bile salt pool, secretion of lithogenic bile	Reduce bowel inflammation; cholecystectomy in symptomatic patients
Primary sclerosing cholangitis (PSC)	Intrahepatic and extrahepatic Inflammation and fibrosis leading to biliary cirrhosis and hepatic failure 7–10% cholangiocarcinoma Small-duct PSC involves small-caliber bile ducts and has a better prognosis	ERCP/high-dose ursodiol lowers risk of colonic neoplasia; cholecystectomy in patients with gallbladder polyps due to the high incidence of malignancy
Urologic		
Nephrolithiasis (10–20%)	CD patients following small-bowel resection; calcium oxalate stones most common	Low-oxalate diet; control of bowel inflammation; surgical intervention
Less common extraintestinal manifestations		
Thromboembolic disorders	Increased risk of venous and arterial thrombosis; factors responsible include abnormalities of the platelet-endothelial interaction, hyperhomocysteinemia, alterations in the coagulation cascade, impaired fibrinolysis, involvement of tissue factor-bearing microvesicles, disruption of the normal coagulation system by autoantibodies, and a genetic predisposition	Anticoagulation; control of inflammation
Cardiopulmonary	Endocarditis, myocarditis, pleuropericarditis, interstitial lung disease	Treatment is varied; stop 5-ASA agents as they can rarely cause interstitial lung disease
Systemic amyloidosis	Secondary (reactive) in long-standing IBD, especially CD	Colchicine
Pancreatitis	Duodenal fistulas, ampullary CD, gallstones, PSC, drugs (MP, azathioprine, 5-ASAs), autoimmune, primary CD of the pancreas	Treatment is varied; stop offending medication; diagnose and treat with ERCP and/or cholecystectomy

Abbreviations: 5-ASA, 5-aminosalicylic acid; DEXA, dual-energy x-ray absorptiometry; ERCP, endoscopic retrograde cholangiopancreatography; IBD, inflammatory bowel disease; IL, interleukin; MP, mercaptopurine; TNF, tumor necrosis factor.

ocular burning. It occurs in 3–4% of IBD patients, more commonly in Crohn's colitis, and is treated with topical glucocorticoids.

■ HEPATOBILIARY

Hepatic steatosis is detectable in about one-half of the abnormal liver biopsies from patients with CD and UC; patients usually present with hepatomegaly. Fatty liver usually results from a combination of chronic debilitating illness, malnutrition, and glucocorticoid therapy. Cholelithiasis occurs in 10–35% of CD patients with ileitis or ileal resection. Gallstone formation is caused by malabsorption of bile acids, resulting in depletion of the bile salt pool and the secretion of lithogenic bile.

Primary sclerosing cholangitis (PSC) is a disorder characterized by both intrahepatic and extrahepatic bile duct inflammation and fibrosis, frequently leading to biliary cirrhosis and hepatic failure; ~5% of patients with UC have PSC, but 50–75% of patients with PSC have IBD. PSC occurs less often in patients with CD. Although it can be recognized after the diagnosis of IBD, PSC can be detected earlier or even years after proctocolectomy. Consistent with this, the immunogenetic basis for PSC appears to be overlapping but distinct from UC based on GWAS, although both IBD and PSC are commonly pANCA positive. Most patients have no symptoms at the time of diagnosis; when symptoms are present, they consist of fatigue, jaundice, abdominal pain, fever, anorexia, and malaise. The traditional gold standard diagnostic test is endoscopic retrograde cholangiopancreatography (ERCP), but magnetic resonance cholangiopancreatography (MRCP) is sensitive, specific, and safer. MRCP is reasonable as an initial diagnostic test in children and adults and can visualize irregularities, multifocal strictures, and dilatations of all levels of the biliary tree. In patients with PSC, both ERCP and MRCP demonstrate multiple bile duct strictures alternating with relatively normal segments.

Gallbladder polyps in patients with PSC have a high incidence of malignancy, and cholecystectomy is recommended, even if a mass lesion is <1 cm in diameter. Gallbladder surveillance with ultrasound should be performed annually. Endoscopic stenting may be palliative for cholestasis secondary to bile duct obstruction. Patients with symptomatic disease develop cirrhosis and liver failure over 5–10 years and eventually require liver transplantation. PSC patients have a 10–15% lifetime risk of developing cholangiocarcinoma and then cannot be transplanted. Patients with IBD and PSC are at increased risk of colon cancer and should be surveyed yearly by colonoscopy and biopsy.

In addition, cholangiography is normal in a small percentage of patients who have a variant of PSC known as *small duct primary sclerosing cholangitis*. This variant (sometimes referred to as “pericholangitis”) is probably a form of PSC involving small-caliber bile ducts. It has similar biochemical and histologic features to classic PSC. It has a significantly better prognosis than classic PSC, although it may evolve into classic PSC. Granulomatous hepatitis and hepatic amyloidosis are much rarer EIMs of IBD.

■ UROLOGIC

The most frequent genitourinary complications are calculi, ureteral obstruction, and ileal bladder fistulas. The highest frequency of nephrolithiasis (10–20%) occurs in patients with CD following small-bowel resection. Calcium oxalate stones develop secondary to hyperoxaluria, which results from increased absorption of dietary oxalate. Normally, dietary calcium combines with luminal oxalate to form insoluble calcium oxalate, which is eliminated in the stool. In patients with ileal dysfunction, however, nonabsorbed fatty acids bind calcium and leave oxalate unbound. The unbound oxalate is then delivered to the colon, where it is readily absorbed, especially in the presence of inflammation.

■ METABOLIC BONE DISORDERS

Low bone mass occurs in 14–42% of IBD patients. The risk is increased by glucocorticoids, CSA, methotrexate (MTX), and total parenteral nutrition (TPN). Malabsorption and inflammation mediated by IL-1, IL-6, TNF, and other inflammatory mediators also contribute to low bone density. An increased incidence of hip, spine, wrist, and rib fractures has been noted: 36% in CD and 45% in UC. The absolute risk of

an osteoporotic fracture is ~1% per person per year. Fracture rates, particularly in the spine and hip, are highest among the elderly (age >60). One study noted an OR of 1.72 for vertebral fracture and an OR of 1.59 for hip fracture. The disease severity predicted the risk of a fracture. Only 13% of IBD patients who had a fracture were on any kind of anti-fracture treatment. Up to 20% of bone mass can be lost per year with chronic glucocorticoid use. The effect is dose-dependent. Budesonide may also suppress the pituitary-adrenal axis and thus carries a risk of causing osteoporosis.

Osteonecrosis is characterized by death of osteocytes and adipocytes and eventual bone collapse. The pain is aggravated by motion and swelling of the joints. It affects the hips more often than knees and shoulders, and in one series, 4.3% of patients developed osteonecrosis within 6 months of starting glucocorticoids. Diagnosis is made by bone scan or MRI, and treatment consists of pain control, cord decompression, osteotomy, and joint replacement.

■ THROMBOEMBOLIC DISORDERS

Patients with IBD have an increased risk of both venous and arterial thrombosis even if the disease is not active. Factors responsible for the hypercoagulable state have included abnormalities of the platelet-endothelial interaction, hyperhomocysteinemia, alterations in the coagulation cascade, impaired fibrinolysis, involvement of tissue factor-bearing microvesicles, disruption of the normal coagulation system by autoantibodies, and a genetic predisposition. A spectrum of vasculitides involving small, medium, and large vessels has also been observed.

■ OTHER DISORDERS

More common cardiopulmonary manifestations include endocarditis, myocarditis, pleuropericarditis, and interstitial lung disease. A secondary or reactive amyloidosis can occur in patients with long-standing IBD, especially in patients with CD. Amyloid material is deposited systemically and can cause diarrhea, constipation, and renal failure. The renal disease can be successfully treated with colchicine. Pancreatitis is a rare EIM of IBD and results from duodenal fistulas; ampullary CD; gallstones; PSC; drugs such as mercaptopurine, azathioprine, or, very rarely, 5-ASA agents; autoimmune pancreatitis; and primary CD of the pancreas.

TREATMENT

Inflammatory Bowel Disease

5-ASA AGENTS

These agents are effective at inducing and maintaining remission in UC. Peroxisome proliferator-activated receptor γ (PPAR- γ) may mediate 5-ASA therapeutic action by decreasing nuclear localization of NF- κ B. Sulfa-free aminosalicylate formulations include alternative azo-bonded carriers, 5-ASA dimers, and delayed-release and controlled-release preparations. Each has the same efficacy as sulfasalazine when equimolar concentrations are used.

Sulfasalazine is effective treatment for mild to moderate UC, but its high rate of side effects limits its use. Although sulfasalazine is more effective at higher doses, at 6 or 8 g/d, up to 30% of patients experience allergic reactions or intolerable side effects such as headache, anorexia, nausea, and vomiting that are attributable to the sulfapyridine moiety. Hypersensitivity reactions, independent of sulfapyridine levels, include rash, fever, hepatitis, agranulocytosis, hypersensitivity pneumonitis, pancreatitis, worsening of colitis, and reversible sperm abnormalities. Sulfasalazine can also impair folate absorption, and patients should be given folic acid supplements.

Balsalazide contains an azo bond binding mesalamine to the carrier molecule 4-aminobenzoyl- β -alanine; it is effective in the colon.

Delzicol and *Asacol HD* (high dose) are enteric-coated forms of mesalamine with the 5-ASA being released at pH >7. They disintegrate with complete breakup of the tablet occurring in many different parts of the gut ranging from the small intestine to the splenic flexure; they have increased gastric residence when taken

with a meal. *Lialda* is a once-a-day formulation of mesalamine (Multi-Matrix System [MMX]) designed to release mesalamine in the colon. The MMX technology incorporates mesalamine into a lipophilic matrix within a hydrophilic matrix encapsulated in a polymer resistant to degradation at a low pH (<7) to delay release throughout the colon. The safety profile appears to be comparable to other 5-ASA formulations.

Apriso is a formulation containing encapsulated mesalamine granules that delivers mesalamine to the terminal ileum and colon via a proprietary extended-release mechanism (Intellicor). The outer coating of this agent (Eudragit L) dissolves at a pH >6. In addition, there is a polymer matrix core that aids in sustained release throughout the colon. Because *Lialda* and *Apriso* are given once daily, an anticipated benefit is improved compliance compared with two to four daily doses required for other mesalamine preparations.

Pentasa is another mesalamine formulation that uses an ethylcellulose coating to allow water absorption into small beads containing the mesalamine. Water dissolves the 5-ASA, which then diffuses out of the bead into the lumen. Disintegration of the capsule occurs in the stomach. The microspheres then disperse throughout the entire GI tract from the small intestine through the distal colon in both fasted and fed conditions.

Salofalk Granu-Stix, an unencapsulated version of mesalamine, has been in use in Europe for induction and maintenance of remission for several years.

Appropriate doses of the 5-ASA compounds are shown in Table 326-8. Some 50–75% of patients with mild to moderate UC improve when treated with 5-ASA doses equivalent to 2 g/d of mesalamine; the dose response continues up to at least 4.8 g/d.

More common side effects of the 5-ASA medications include headaches, nausea, hair loss, and abdominal pain. Rare side effects of the 5-ASA medications include renal impairment, hematuria, pancreatitis, and paradoxical worsening of colitis. Renal function tests and urinalysis should be checked yearly.

Topical *Rowase* enemas are composed of mesalamine and are effective in mild-to-moderate distal UC. Combination therapy with mesalamine in both oral and enema form is more effective than either treatment alone for both distal and extensive UC.

Canasa suppositories composed of mesalamine are effective in treating proctitis.

GLUCOCORTICOIDS

The majority of patients with moderate to severe UC benefit from oral or parenteral glucocorticoids. Prednisone is usually started at doses of 40–60 mg/d for active UC that is unresponsive to 5-ASA therapy. Parenteral glucocorticoids may be administered as hydrocortisone, 300 mg/d, or methylprednisolone, 40–60 mg/d. A newer glucocorticoid for UC, budesonide (Uceris), is released entirely in

the colon and has minimal to no glucocorticoid side effects. The dose is 9 mg/d for 8 weeks, and no taper is required. Topically applied glucocorticoids (hydrocortisone enemas or budesonide foam) are also beneficial for distal colitis and may serve as an adjunct in those who have rectal involvement plus more proximal disease. Hydrocortisone enemas are significantly absorbed from the rectum and can lead to adrenal suppression with prolonged administration. Topical 5-ASA therapy is more effective than topical steroid therapy in the treatment of distal UC.

Glucocorticoids are also effective for treatment of moderate to severe CD and induce a 60–70% remission rate compared to a 30% placebo response. The systemic effects of standard glucocorticoid formulations have led to the development of formulations that are less well absorbed and have increased first-pass metabolism. Controlled-ileal-release budesonide has been nearly equal to prednisone for ileocolonic CD with fewer glucocorticoid side effects. Budesonide is used for 2–3 months at a dose of 9 mg/d and then tapered. Glucocorticoids play no role in maintenance therapy in either UC or CD. Once clinical remission has been induced, they should be tapered according to the clinical activity, normally at a rate of no more than 5–10 mg/week. The side effects are numerous, including fluid retention, abdominal striae, fat redistribution, hyperglycemia, subcapsular cataracts, osteonecrosis, osteoporosis, myopathy, emotional disturbances, and withdrawal symptoms. Most of these side effects, aside from osteonecrosis, are related to the dose and duration of therapy.

ANTIBIOTICS

Antibiotics have no role in the treatment of active or quiescent UC. However, pouchitis, which occurs in ~30–50% of UC patients after colectomy and IPAA, usually responds to treatment with a variety of antibiotics including metronidazole and ciprofloxacin. Some patients require long-term treatment with antibiotics for chronic pouchitis.

AZATHIOPRINE AND MERCAPTOPURINE

Azathioprine and mercaptopurine (MP) are purine analogues used concomitantly with biologic therapy or, much less often, as the sole immunosuppressants. Azathioprine is rapidly absorbed and converted to MP, which is then metabolized to the active end product, thioguanine, an inhibitor of purine ribonucleotide synthesis and cell proliferation. Efficacy can be seen as early as 3–4 weeks but can take up to 4–6 months. Adherence can be monitored by measuring the levels of 6-thioguanine and 6-methylmercaptopurine, end products of MP metabolism. The doses used range from 2 to 3 mg/kg per day for azathioprine and 1 to 1.5 mg/kg per day for MP.

Although azathioprine and MP are usually safe, pancreatitis occurs in 3–4% of patients, typically presents within the first few weeks of therapy, and is completely reversible when the drug is

TABLE 326-8 Oral 5-Aminosalicylic Acid (5-ASA) Preparations

PREPARATION	FORMULATION	DELIVERY	DOSING PER DAY
Azo-Bond			
Sulfasalazine (500 mg) (Azulfidine)	Sulfapyridine-5-ASA	Colon	3–6 g (acute) 2–4 g (maintenance)
Balsalazide (750 mg) (Colazal)	Aminobenzoyl-alanine-5-ASA	Colon	6.75–9 g
Delayed-Release			
Mesalamine (400, 800 mg) (Delzicol, Asacol HD)	Eudragit S (pH 7)	Distal ileum-colon	2.4–4.8 g (acute) 1.6–4.8 g (maintenance)
Mesalamine (1.2 g) (Lialda)	MMX mesalamine (SPD476)	Ileum-colon	2.4–4.8 g
Controlled-Release			
Mesalamine (250, 500, 1000 mg) (Pentasa)	Ethylcellulose microgranules	Stomach-colon	2–4 g (acute) 1.5–4 g (maintenance)
Delayed- and Extended-Release			
Mesalamine (0.375 g) (Apriso)	Intellicor extended-release mechanism	Ileum-colon	1.5 g (maintenance)

Abbreviation: MMX, Multi-Matrix System.

stopped. Other side effects include nausea, fever, rash, and hepatitis. Bone marrow suppression (particularly leukopenia) is dose-related and often delayed, necessitating regular monitoring of the complete blood cell count (CBC). Additionally, 1 in 300 individuals lacks thiopurine methyltransferase, the enzyme responsible for drug metabolism to inactive end products (6-methylmercaptopurine); an additional 11% of the population are heterozygotes with intermediate enzyme activity. Both are at increased risk of toxicity because of increased accumulation of active 6-thioguanine metabolites. Although 6-thioguanine and 6-methylmercaptopurine levels can be followed to determine correct drug dosing and reduce toxicity, weight-based dosing is an acceptable alternative. CBCs and liver function tests should be monitored frequently regardless of dosing strategy.

One meta-analysis demonstrated a fourfold risk of lymphoma in IBD patients on azathioprine and MP. The highest risk for thiopurine-associated lymphoma is in patients >65 years old actively using thiopurines (yearly incidence rate per 1000 patient-years of 5.41), with a moderate risk in those between the ages of 50 and 65 (incidence rate of 2.58 compared to an incidence rate of 0.37 in patients <50 years old). Patients using thiopurines also have a two- to threefold increased risk of nonmelanoma skin cancers.

METHOTREXATE

MTX inhibits dihydrofolate reductase, resulting in impaired DNA synthesis. Additional anti-inflammatory properties may be related to decreases in the production of IL-1. It is used most often concomitantly with biologic therapy to decrease antibody formation and improve disease response. Intramuscular (IM) or subcutaneous (SC) doses range from 15 to 25 mg/week. Potential toxicities include leukopenia and hepatic fibrosis, necessitating periodic evaluation of CBCs and liver enzymes. The role of liver biopsy in patients on long-term MTX is uncertain but is probably limited to those with increased liver enzymes. Hypersensitivity pneumonitis is a rare but serious complication of therapy.

CYCLOSPORINE

CSA is a lipophilic peptide with inhibitory effects on both the cellular and humoral immune systems. CSA blocks the production of IL-2 by T helper lymphocytes. CSA binds to cyclophilin, and this complex inhibits calcineurin, a cytoplasmic phosphatase enzyme involved in the activation of T cells. CSA also indirectly inhibits B-cell function by blocking helper T cells. CSA has a more rapid onset of action than MP and azathioprine.

CSA is most effective when given at 2–4 mg/kg per day IV in severe UC that is refractory to IV glucocorticoids, with 82% of patients responding. CSA can be an alternative to colectomy. The long-term success of oral CSA is not as dramatic, but if patients are started on MP or azathioprine at the time of hospital discharge, remission can be maintained. Levels as measured by monoclonal radioimmunoassay or by the high-performance liquid chromatography assay should be maintained between 150 and 350 ng/mL.

CSA may cause significant toxicity; renal function should be monitored frequently. Hypertension, gingival hyperplasia, hypertrichosis, paresthesias, tremors, headaches, and electrolyte abnormalities are common side effects. Creatinine elevation calls for dose reduction or discontinuation. Seizures may also complicate therapy, especially if the patient is hypomagnesemic or if serum cholesterol levels are <3.1 mmol/L (<120 mg/dL). Opportunistic infections, most notably *Pneumocystis jirovecii* pneumonia, may occur with combination immunosuppressive treatment; antibiotic prophylaxis with trimethoprim-sulfamethoxazole should be given.

To compare IV CSA versus infliximab, a large trial was conducted in Europe by the GETAID (Group d'Etudes Thérapeutiques des Affections Inflammatoires Digestives) group. The results indicated identical 7-day response rates for CSA 2 mg/kg (with doses adjusted for levels of 150–250 ng/mL) and infliximab 5 mg/kg, with both groups achieving response rates of 85%. Serious infections occurred in 5 of 55 CSA patients and 4 of 56 infliximab patients. Response rates were similar in the two groups at day 98

among patients treated with oral CSA versus infliximab at the usual induction dose and maintenance dose regimen (40 and 46%, respectively). In light of data showing equal efficacy of CSA and infliximab in severe UC, more physicians are relying on infliximab rather than CSA in these patients.

TACROLIMUS

Tacrolimus is a macrolide antibiotic with immunomodulatory properties similar to CSA but 100 times as potent and not dependent on bile or mucosal integrity for absorption. Thus, tacrolimus has good oral absorption despite proximal small-bowel Crohn's involvement. Tacrolimus is effective in children with refractory IBD and in adults with extensive involvement of the small bowel. It is also effective in adults with glucocorticoid-dependent or refractory UC and CD as well as refractory fistulizing CD.

BIOLOGIC THERAPIES

Biologic therapy is now commonly given as an initial therapy for patients with moderate to severe CD and UC to prevent future complications of IBD. High-risk patients with UC who are more likely to require biologics include those with moderate to severe disease, steroid-dependent or steroid-refractory disease, and refractory pouchitis. High-risk patients with CD who are more likely to require biologics include those who are <30 years old, with extensive disease, perianal or severe rectal disease and/or deep ulcerations in the colon, and stricturing or penetrating disease behavior. The current goal of IBD treatment is to treat early in the disease course, treat aggressively with biologics, check drug and drug metabolite levels, administer dual therapy with immunomodulators and biologics in appropriate patients, and aim for deep remission (endoscopic and histologic remission). Patients who respond to biologic therapies enjoy an improvement in clinical symptoms; a better quality of life; less disability, fatigue, and depression; and fewer surgeries and hospitalizations.

Anti-TNF Therapies TNF is a proinflammatory cytokine that regulates immune cells to coordinate a systemic immune response. Dysregulation of TNF production has been associated with immune-mediated disorders including IBD, and inhibition of TNF signaling is used in the treatment of IBD. Four TNF inhibitors are currently approved for the treatment of IBD: infliximab, adalimumab, certolizumab pegol, and golimumab. Infliximab, a chimeric IgG1 antibody against TNF- α , was the first biologic therapy approved for moderately to severely active inflammatory and fistulizing CD and UC.

The SONIC (Study of Biologic and Immunomodulator-Naive Patients with Crohn's Disease) trial compared infliximab plus azathioprine, infliximab alone, and azathioprine alone in immunomodulator- and biologic therapy-naïve patients with moderate to severe CD. At 1 year, the infliximab plus azathioprine group had a glucocorticoid-free remission rate of 46% compared with 35% for infliximab alone and 24% for azathioprine alone. Complete mucosal healing was noted in more patients at week 26 with the combined approach compared with either infliximab or azathioprine alone (44 vs 30 vs 17%). The adverse events were equal between groups.

A similar study in patients with moderate to severe UC showed that after 16 weeks of therapy, UC patients receiving azathioprine plus infliximab exhibited a glucocorticoid-free remission rate of 40%, compared to rates of 24 and 22% in those on azathioprine and infliximab alone, respectively. Together, these studies support a more aggressive therapy for moderate to severe CD and UC. Trough infliximab levels can be checked, and if low, the dose can be increased or the interval decreased.

Hospitalized patients with acute severe glucocorticoid-refractory UC have a high inflammatory burden and may develop a protein-losing enteropathy, leading to an accelerated consumption, excessive fecal wasting, and low serum concentrations of infliximab. Given a clear exposure-response relationship for infliximab in patients with IBD, intensive infliximab dosing regimens have been used in these patients.

Adalimumab (ADA) is a recombinant human monoclonal IgG1 antibody containing only human peptide sequences and is injected subcutaneously. ADA binds TNF and neutralizes its function by blocking the interaction between TNF and its cell-surface receptor. Therefore, it seems to have a similar mechanism of action to infliximab but with less immunogenicity. ADA is approved for treatment of moderate to severe CD and UC. CHARM (Crohn's Trial of the Fully Human Adalimumab for Remission Maintenance) is an ADA maintenance study in patients who responded to ADA induction therapy. About 50% of the patients in this trial were previously treated with infliximab. Remission rates ranged from 42 to 48% in infliximab-naïve patients at 1 year compared with remission rates of 31–34% in patients who had previously received infliximab. UC results are similar, with a sustained remission rate at 1 year of 22% (12.4% placebo) among anti-TNF-naïve patients and a sustained remission rate at 1 year of 10.2% (3% placebo) among patients who had previously received anti-TNF agents. In clinical practice, the remission rate in both CD and UC patients taking ADA increases with a dose increase to 40 mg weekly instead of every other week.

Certolizumab pegol is a pegylated form of an anti-TNF Fab portion of an antibody administered SC once monthly. SC certolizumab pegol was effective for induction of clinical response in patients with active inflammatory CD.

Golimumab is another fully human IgG1 antibody against TNF- α and is currently approved for the treatment of moderately to severely active UC. Like ADA and certolizumab, golimumab is injected SC.

Side Effects of Anti-TNF Therapies

Development of Antibodies and Drug Levels The development of antibodies to infliximab is associated with an increased risk of infusion reactions and a decreased response to treatment. Current practice does not include giving on-demand or episodic infusions in contrast to scheduled periodic infusions because patients are most likely to develop antibodies. Anti-infliximab antibodies are generally present when the quality of response or the response duration to infliximab infusion decreases. Commercial assays can detect both infliximab and ADA antibodies and measure trough levels to determine optimal dosing. If a patient has high anti-infliximab antibodies and a low trough level of infliximab, it is best to switch to another anti-TNF therapy. If a patient has a therapeutic anti-TNF level and active inflammatory symptoms, the drug should be switched to a different class of biologic. Most acute infusion reactions and serum sickness can be managed with glucocorticoids and antihistamines. Some reactions can be serious and would necessitate a change in therapy, especially if a patient has anti-infliximab antibodies. It is now common practice to add an immunomodulator such as azathioprine, MP, or MTX to anti-TNF therapy to help prevent antibody formation.

Non-Hodgkin's Lymphoma (NHL) The baseline risk of NHL in CD patients is 2 in 10,000, slightly higher than in the general population. Azathioprine and/or MP therapy increases the risk to ~4 in 10,000. It is difficult to assess whether anti-TNF medications are associated with lymphoma because most patients are also receiving thiopurines. After adjustment for co-treatments, no excess risk of lymphoma was found in a Danish study of a cohort of IBD patients exposed to anti-TNF medications.

Hepatosplenic T-Cell Lymphoma (HSTCL) HSTCL is a nearly universally fatal lymphoma in patients with or without CD. In patients with CD, a total of 37 unique cases have been reported. Eighty-six percent of the patients were male, and the median age was 26 years. Patients had CD for a mean of 10 years before the diagnosis of HSTCL. Thirty-six patients had used either MP or azathioprine, and 28 patients had used infliximab.

Skin Lesions New-onset psoriasiform skin lesions develop in nearly 5% of IBD patients treated with anti-TNF therapy. Most often, these can be treated topically, and occasionally, anti-TNF therapy must be decreased, switched, or stopped. Patients with IBD

may have a slight, unexplained, intrinsic higher risk of developing melanoma. The risk of melanoma is increased almost twofold with anti-TNF and not thiopurine use. The risk of nonmelanoma skin cancer is increased with thiopurines and biologics, especially with ≥1 year of follow-up. Patients on these medications should have a skin check at least once a year.

Infections All of the anti-TNF drugs are associated with an increased risk of infections, particularly reactivation of latent tuberculosis and opportunistic fungal infections including disseminated histoplasmosis and coccidioidomycosis. Patients should have a purified protein derivative (PPD) or a QuantiFERON-TB Gold test before initiation of anti-TNF therapy. Patients >65 years old have a higher rate of infections and death on infliximab or ADA than those <65 years old.

Other Acute liver injury due to reactivation of hepatitis B virus and to autoimmune effects and cholestasis has been reported. Rarely, infliximab and the other anti-TNF drugs have been associated with optic neuritis, seizures, new onset or exacerbation of clinical symptoms, and radiographic evidence of central nervous system demyelinating disorders, including multiple sclerosis. They may exacerbate symptoms in patients with New York Heart Association functional class III/IV heart failure.

ANTI-INTEGRINS

Integrins are expressed on the cell surface of leukocytes and serve as mediators of leukocyte adhesion to vascular endothelium. $\alpha 4$ -Integrin along with its $\beta 1$ or $\beta 7$ subunit interact with endothelial ligands termed adhesion molecules. Interaction between $\alpha 4\beta 7$ and mucosal addressin cellular adhesion molecule (MAdCAM-1) is important in lymphocyte trafficking to gut mucosa.

Natalizumab is a recombinant humanized IgG4 antibody against $\alpha 4$ -integrin and is effective in induction and maintenance of patients with CD. The rates of response and remission at 3 months are ~60 and 40%, respectively, with a sustained remission rate of ~40% at 36 weeks. Natalizumab is no longer widely used for CD due to the risk of progressive multifocal leukoencephalopathy (PML).

Vedolizumab (VDZ), another leukocyte trafficking inhibitor, is a monoclonal antibody directed against $\alpha 4\beta 7$ -integrin specifically and has the ability to convey gut-selective immunosuppression. Unlike natalizumab, it inhibits adhesion of a discrete gut-homing subset of T lymphocytes to MAdCAM-1, but not to vascular adhesion molecule-1. VDZ decreases GI inflammation without inhibiting systemic immune responses or affecting T-cell trafficking to the central nervous system. It may be prescribed as a first-line biologic or after failure of a TNF antagonist in patients with CD or UC. The VARSITY trial, a phase 3B, randomized, double-blind, double-dummy, active-controlled superiority trial, evaluated outcomes among patients with UC who received either VDZ or ADA. Results showed that, at week 52, patients who were treated with VDZ were significantly more likely to be in clinical remission (31.3% VDZ vs 22.5% ADA) and show endoscopic improvement (39.7% VDZ vs 27.7% ADA). Glucocorticoid-free clinical remission was observed in 12.6% of the VDZ group and 21.8% of patients who received ADA, but the difference was not statistically significant. This trial suggests that among patients with UC, VDZ should be considered as first-line therapy and before treatment with ADA.

Ustekinumab, a fully human IgG1 monoclonal antibody, blocks the biologic activity of IL-12 and IL-23 through their common p40 subunit by inhibiting the interaction of these cytokines with their receptors on T cells, natural killer cells, and antigen-presenting cells. In the UNITI trial, the remission rate for the highest 6 mg/kg IV induction dose followed by a dose of 90 mg every 8 weeks was 41.7%, compared with 27.4% for placebo, at 22 weeks in patients with CD no longer responding to anti-TNF therapy.

Similarly, the UNIFI trial evaluated ustekinumab as 8-week induction and 44-week maintenance therapy in moderate to severe UC. Induction rates at 8 weeks were 15.6% in the ustekinumab group compared to 5.3% in the placebo group, and 44-week maintenance rates were 43.8% in the ustekinumab group compared to

24% in the placebo group. The rates of serious adverse events were similar for ustekinumab and placebo in the UNITI and UNIFI trials. Therefore, ustekinumab is another option for the treatment of moderate to severe CD and UC and is particularly appealing for use in patients with concomitant psoriatic arthritis.

SMALL MOLECULES

Small molecules (drugs with molecular weight <1 kDa) are a new class of orally administered medications developed for IBD that lack the immunogenicity associated with monoclonal antibodies. The advantage of small molecules is their ability to diffuse through cell membranes into the intracellular space and alter cytokine signaling pathways. This mechanism of action may be more efficacious compared to monoclonal antibodies that inhibit specific targets because several cytokine pathways are involved in IBD pathogenesis and inhibiting numerous cytokines may be synergistic. A key regulatory pathway is the JAK/STAT pathway that activates transcription and translation of proteins that mediate the immune response. Janus kinase (JAK) is a family of intracellular, nonreceptor tyrosine kinases that regulate cytokine signaling via the JAK/STAT pathway, ultimately suppressing the immune response and inflammation. The JAK family members include JAK1, JAK2, JAK3, and tyrosine kinase 2 (TYK2).

Tofacitinib is a reversible and competitive JAK inhibitor used for the treatment of moderate to severe UC refractory to conventional therapy. It competes with ATP to bind to the ATP-docking site of the kinase domain of JAK. By competing with ATP, tofacitinib inhibits phosphorylation and activation of JAK, leading to downstream reduction of cytokine production and alteration of the immune response. Although tofacitinib is a pan-JAK inhibitor, it has higher specificity for JAK1 and JAK3 than for JAK2 and TYK2. The pan-JAK inhibition is concerning for adverse events and overall safety.

The efficacy of tofacitinib as induction and maintenance therapy, as well as its safety profile, was evaluated in three phase 3, randomized, double-blind, placebo-controlled trials in adults with moderate to severe UC refractory to conventional therapy including anti-TNFs. Patients who responded to induction therapy were eligible for OCTAVE Sustain, a maintenance trial of tofacitinib 5 mg versus 10 mg versus placebo that continued through 52 weeks, with the primary end point of clinical remission at 52 weeks. Remission rates at 8 weeks in the OCTAVE Induction 1 and 2 trials were 18.5 and 16.6% in the tofacitinib groups, compared to 8.2 and 3.6% in the placebo groups, respectively. In the OCTAVE Sustain trial, remission rates at 52 weeks were 34.3% with 5 mg and 40.6% with 10 mg of tofacitinib, compared to 11.1% with placebo. A recent U.S. Food and Drug Administration review concluded that there is an increased risk of serious adverse events including heart attack, stroke, cancer, blood clots, and death in patients with ulcerative colitis and rheumatoid arthritis who are prescribed tofacitinib. Patients who are at risk for cardiovascular disease, are current or past smokers and/or are over the age of 50 should consider alternative therapies.

OZANIMOD

Ozanimod is a potent sphingosine-1-phosphate (SIP1) receptor modulator that binds selectively with high affinity to the SIP receptor subtypes SIP1 and SIP5, both of which are involved in immune regulation. By preventing trafficking of disease-exacerbating lymphocytes to the gut, ozanimod may provide immunomodulatory effects and moderate disease processes.

Ozanimod has very recently been approved for the treatment of moderate to severe ulcerative colitis. It is administered as a daily capsule.

The biologic and small-molecule therapies used in daily practice are detailed in [Table 326-9](#).

NUTRITIONAL THERAPIES

Diet has long been thought to contribute to the pathogenesis of IBD and may also be an avenue for managing disease activity. Diet plays a significant role in shaping the gut microbiome, and dietary components may interact with the microbiome and stimulate a mucosal immune response. In fact, active CD responds to exclusive enteral nutrition (EEN) or bowel rest with TPN, interventions as effective

as glucocorticoids in inducing remission but not as effective for maintenance therapy. In contrast to CD, active UC is not effectively treated by elemental diets or TPN.

Dietary approaches to maintenance therapy in CD have largely been adapted from epidemiologic studies; however, significant heterogeneity is noted among research study outcomes. In general, low fiber, refined carbohydrates (especially sweetened beverages), animal fats, red meat, and processed meat have been associated with onset of IBD. Therefore, the overall dietary approach is to maximize fiber intake, particularly from fruits and vegetables, and to limit consumption of higher-risk foods. Several defined diets adhere to these principles with some variation, including the Mediterranean diet pattern, specific carbohydrate diet, semi-vegetarian diet, and IBD anti-inflammatory diet (IBD-AID). However, it remains unclear whether diet studies will eventually lead to evidence-based nutrition guidelines.

Standard medical management of UC and CD is shown in [Fig. 326-12](#).

SURGICAL THERAPY

Ulcerative Colitis Nearly one-half of patients with extensive chronic UC undergo surgery within the first 10 years of their illness. The indications for surgery are listed in [Table 326-10](#). Morbidity is ~20% for elective, 30% for urgent, and 40% for emergency proctocolectomy. The risks are primarily hemorrhage, contamination and sepsis, and neural injury. The operation of choice is an IPAA.

Because UC is a mucosal disease, the rectal mucosa can be dissected and removed down to the dentate line of the anus or ~2 cm proximal to this landmark. The ileum is fashioned into a pouch that serves as a neorectum. This ileal pouch is then sutured circumferentially to the anus in an end-to-end fashion. If performed carefully, this operation preserves the anal sphincter and maintains continence. The overall operative morbidity is 10%, with the major complication being bowel obstruction. Pouch failure necessitating conversion to permanent ileostomy occurs in 5–10% of patients. Some inflamed rectal mucosa is usually left behind, and thus, endoscopic surveillance is necessary. Primary dysplasia of the ileal mucosa of the pouch has occurred rarely.

Patients with IPAA usually have ~6–10 bowel movements a day. On validated quality-of-life indices, they report better performance in sports and sexual activities than ileostomy patients. The most frequent complication of IPAA is pouchitis in ~30–50% of patients with UC. This syndrome consists of increased stool frequency, watery stools, cramping, urgency, nocturnal leakage of stool, arthralgias, malaise, and fever. Pouch biopsies may distinguish true pouchitis from underlying CD. Although pouchitis usually responds to antibiotics, 3–5% of patients remain refractory and may require glucocorticoids, immunomodulators, biologics, or even pouch removal.

Crohn's Disease The majority of patients with CD will require at least one operation in their lifetime. The need for surgery is related to duration of disease and the site of involvement. Patients with small-bowel disease have an 80% chance of requiring surgery. Those with colitis alone have a 50% chance. Surgery is an option only when medical treatment has failed or complications dictate its necessity. The indications for surgery are shown in [Table 326-10](#).

Small-Intestinal Disease Because CD is chronic and recurrent, with no clear surgical cure, as little intestine as possible is resected. Current surgical alternatives for treatment of obstructing CD include resection of the diseased segment and strictureplasty. Surgical resection of the diseased segment is the most frequently performed operation, and in most cases, primary anastomosis can be done to restore continuity. An end-to-end anastomosis may provide the best opportunity for an optimal functional outcome, compared to an antiperistaltic side-to-side anastomosis, which creates a functional block to motility leading to distention and pain at the anastomotic site in a subgroup of patients. If much of the small bowel has already been resected and the strictures are short, with intervening areas of normal mucosa, strictureplastics should be done to avoid a functionally insufficient length of bowel. The

TABLE 326-9 Biologic Agents in the Treatment of Inflammatory Bowel Disease

MEDICATION	DOSAGE	INDICATION	SERIOUS TOXICITIES	OTHER COMMON SIDE EFFECTS	TESTING
Infliximab	5 mg/kg at 0, 2, and 6 weeks; then every 8 weeks; may increase dose to 10 mg/kg every 4 weeks depending on trough levels Intensive dosing for hospitalized corticosteroid-refractory patients	Moderate to severe Crohn's disease and ulcerative colitis Fistulizing Crohn's disease	Increased risk of infections (bacterial and fungal), tuberculosis (TB) reactivation, hepatitis B reactivation, lymphoma (controversial), psoriasis, melanoma and nonmelanoma skin cancers, drug-induced lupus Contraindicated in multiple sclerosis, class III/IV congestive heart failure	Infusion reactions	Prior to infusion: TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Skin check yearly Influenza, Pneumovax 23, and Prevnar 13 vaccinations Hepatitis B vaccine if not immune
Adalimumab	160 mg day 0, 80 mg day 14 and then 40 mg every 14 days; may increase to 40 mg every 7 days depending on trough levels	Moderate to severe Crohn's disease and ulcerative colitis. Fistulizing Crohn's disease	As above	Injection site reactions (better with citrate-free preparation)	As above
Certolizumab	400 mg on days 0 and 14, then 400 mg every 28 days	Moderate to severe Crohn's disease	As above	As above	As above
Golimumab	200 mg on day 0, 100 mg on day 14, then 100 mg every 28 days	Moderate to severe ulcerative colitis	As above	As above	As above
Vedolizumab	300 mg at 0, 2, and 6 weeks, then every 8 weeks; may increase dose to 300 mg every 4 weeks	Moderate to severe ulcerative colitis (more effective than adalimumab as first-line therapy in one study)	No increased risk of serious systemic or opportunistic infections No increased risk of malignancy	Nasopharyngitis, headache, arthralgias, nausea	Prior to infusion: TB testing hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, Pneumovax 23, and Prevnar 13 vaccinations Hepatitis B vaccine if not immune
Natalizumab	300 mg IV every 4 weeks	Moderate to severe Crohn's disease (not to be used in combination with other immunosuppressive medications)	Progressive multifocal leukoencephalopathy (monitor anti-JCV antibodies every 6 months and stop if positive)	Headache, fatigue, infusions reactions, urinary tract infections, arthralgia, pain in extremity, rash, gastroenteritis, vaginitis	Prior to infusion: Anti-JCV antibody, TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, Pneumovax 23, and Prevnar 13 vaccinations Hepatitis B vaccine if not immune
Ustekinumab	6 mg/kg IV, then 90 mg every 8 weeks; may increase dose to 90 mg every 4 weeks	Moderate to severe Crohn's disease and ulcerative colitis	Reversible posterior leukoencephalopathy syndrome (presents with headaches, seizures, confusion, and visual disturbances), anaphylaxis, and angioedema	Nasopharyngitis, upper respiratory tract infection, fatigue, headache	Prior to infusion: TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, Pneumovax 23, and Prevnar 13 vaccinations Hepatitis B vaccine if not immune
Tofacitinib	10 mg bid; can decrease to 5 mg bid when patient in remission	Moderate to severe ulcerative colitis	Increased risk of heart attack, stroke, cancer, blood clots, and death in patients with ulcerative colitis and rheumatoid arthritis. Patients who are at risk for cardiovascular disease, are current or past smokers and/or are over the age of 50 should consider alternative therapies. Increased risk of viral infections, including herpes zoster, and bacterial and invasive fungal infections	Elevated lipids, neutropenia, anemia, elevated liver enzymes	Prior to infusion: First dose of Shingrix recommended, TB testing Hepatitis B testing (HBsAb, HBsAg, HBcAb) Maintenance: Influenza, Pneumovax 23, and Prevnar 13 vaccinations Hepatitis B vaccine if not immune

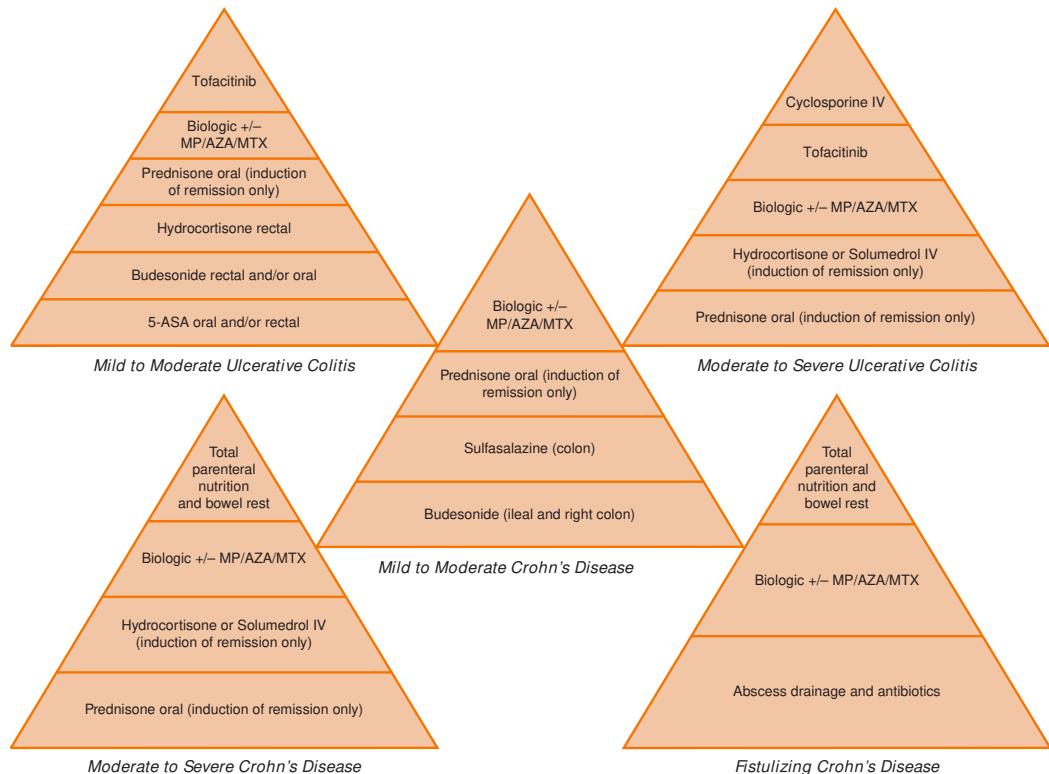


FIGURE 326-12 Medical management of inflammatory bowel disease. 5-ASA, 5-aminosalicylic acid; CD, Crohn's disease; UC, ulcerative colitis.

strictured area of intestine is incised longitudinally and the incision sutured transversely, thus widening the narrowed area. Complications of strictureplasty include prolonged ileus, hemorrhage, fistula, abscess, leak, and restricture.

Risk factors for early recurrence of disease include cigarette smoking, penetrating disease (internal fistulas, abscesses, or other evidence of penetration through the wall of the bowel), early recurrence since a previous surgery, multiple surgeries, and a young age at the time of the first surgery. Aggressive postoperative treatment with biologics should be considered for this group of patients. It is also recommended to evaluate for endoscopic recurrence of CD via a colonoscopy, if possible, 3–6 months after surgery.

Colorectal Disease A greater percentage of patients with Crohn's colitis require surgery for intractability, fulminant disease, and anorectal disease. Several alternatives are available, ranging

from the use of a temporary loop ileostomy to resection of segments of diseased colon or even the entire colon and rectum. For patients with segmental involvement, segmental colon resection with primary anastomosis can be performed. In 20–25% of patients with extensive colitis, the rectum is spared sufficiently to consider rectal preservation. Most surgeons believe that an IPAA is contraindicated in CD due to the high incidence of pouch failure. A diverting colostomy may help heal severe perianal disease or rectovaginal fistulas, but disease almost always recurs with reanastomosis. These patients often require a total proctocolectomy and ileostomy.

IBD AND PREGNANCY

Patients with quiescent UC and CD have normal fertility rates; the fallopian tubes can be scarred by the inflammatory process of CD, especially on the right side because of the proximity of the terminal ileum. In addition, perirectal, perineal, and rectovaginal abscesses and fistulas as well as pelvic surgery can result in dyspareunia. Infertility in men can be caused by sulfasalazine but reverses when treatment is stopped. Women with an IPAA have decreased fertility due to scarring or occlusion of the fallopian tubes secondary to pelvic inflammation and adhesions, although studies have shown that fertility is improved with laparoscopic versus open IPAA.

Mild or quiescent UC or CD has no effect on birth outcomes. The courses of CD and UC during pregnancy mostly correlate with disease activity at the time of conception. Patients should be in remission for 6 months before conceiving. Most CD patients can deliver vaginally, but cesarean delivery may be the preferred route of delivery for patients with anorectal and perirectal abscesses and fistulas to reduce the likelihood of fistulas developing or extending into the episiotomy scar. Unless they desire multiple children, UC patients with an IPAA may consider a cesarean delivery due to an increased risk of future fecal incontinence.

Sulfasalazine and all mesalamines are safe for use in pregnancy and nursing with the caveat that additional folate supplementation must be given with sulfasalazine. Topical 5-ASA agents are safe during pregnancy and nursing. Glucocorticoids are generally safe for use during pregnancy and are indicated for patients with moderate to severe

TABLE 326-10 Indications for Surgery

ULCERATIVE COLITIS	CROHN'S DISEASE
Intractable disease	Small Intestine
Fulminant disease	Stricture and obstruction unresponsive to medical therapy
Toxic megacolon	Massive hemorrhage
Colonic perforation	Refractory fistula
Massive colonic hemorrhage	Abscess
Extracolonic disease	Colon and rectum
Colonic obstruction	Intractable disease
Colon cancer prophylaxis	Fulminant disease
Colon dysplasia or cancer	Perianal disease unresponsive to medical therapy
	Refractory fistula
	Colonic obstruction
	Cancer prophylaxis
	Colon dysplasia or cancer

disease activity. The amount of glucocorticoids received by the nursing infant is minimal. The safest antibiotics to use for CD in pregnancy for short periods of time (weeks, not months) are ampicillin and cephalosporins. Metronidazole can be used in the second or third trimester. Ciprofloxacin causes cartilage lesions in immature animals and should be avoided because of the absence of data on its effects on growth and development in humans.

MP and azathioprine pose minimal or no risk during pregnancy. Breast milk has been shown to contain negligible levels of MP/azathioprine when measured in a limited number of patients.

MTX is teratogenic and should be discontinued at least 3 months before conception.

In a large prospective and multiple retrospective studies, no increased risk of stillbirths, miscarriages, or spontaneous abortions was seen with infliximab, ADA, or certolizumab. Infliximab and ADA are IgG1 antibodies and are actively transported across the placenta in the late second and third trimesters. Infants can have serum levels of infliximab and ADA up to 12 months of age, and live vaccines should be avoided during this time. Certolizumab crosses the placenta by passive diffusion, and infant serum and cord blood levels are minimal. The anti-TNF drugs are relatively safe in nursing. Minuscule levels of infliximab, ADA, and certolizumab have been reported in breast milk. These levels are of no clinical significance. It is recommended that drugs should not be switched during pregnancy unless necessitated by the medical condition of the IBD. VDZ and ustekinumab appear safe during pregnancy, although the data are limited. Tofacitinib should not be used during pregnancy. Animal studies show teratogenic effects with tofacitinib, and data in humans are limited. A washout period of at least 1 week is recommended before conception. Surgery in UC should be performed only for emergency indications, including severe hemorrhage, perforation, and megacolon refractory to medical therapy. Total colectomy and ileostomy carry a 50% risk of postoperative spontaneous abortion. The best time to perform surgery is in the second trimester if necessary. Patients with IPAAAs have increased nighttime stool frequency during pregnancy that resolves postpartum. Transient small-bowel obstruction or ileus has been noted in up to 8% of patients with ileostomies.

CANCER IN IBD

■ ULCERATIVE COLITIS

Patients with long-standing UC are at increased risk for developing colonic epithelial dysplasia and carcinoma (Fig. 326-13).

The risk of neoplasia in chronic UC increases with duration and extent of disease. In contrast to the relatively high risk in one large meta-analysis (2% after 10 years, 8% after 20 years, and 18% after 30 years of disease), a decrease in the risk of colorectal cancer has been noted over time potentially due to better control of inflammation and better colonoscopic surveillance. The rates of colon cancer are still

about 1.5 to 2 times higher than in the general population, and colonoscopic surveillance is the standard of care.

Annual or biennial colonoscopy with multiple biopsies is recommended for patients with >8–10 years of extensive colitis (greater than one-third of the colon involved) or 12–15 years of proctosigmoiditis (less than one-third but more than just the rectum) and has been widely used to screen and survey for subsequent dysplasia and carcinoma. International guideline societies have recommended chromoendoscopy for dysplasia surveillance in IBD. Chromoendoscopy enhances the visualization of the surface and pit pattern of the mucosa, as well as borders of lesions, in order to better define areas of dysplasia compared to standard-definition white light endoscopy. The evidence behind chromoendoscopy is controversial. A systematic review of randomized controlled trials found that chromoendoscopy had a higher likelihood of detecting dysplasia compared to standard-definition white light endoscopy with a relative risk of 2.12. In contrast, a retrospective study found no significant difference in dysplasia detection rates between chromoendoscopy and standard-definition white light endoscopy. In real-life settings, the practice has been to use standard-definition white light endoscopy with surveillance biopsies in patients with chronic colitis at average risk and chromoendoscopy in higher-risk patients including those with a history of dysplasia, PSC, or family history of colorectal cancer.

Risk factors for cancer in UC include long-duration disease, extensive disease, family history of colon cancer, PSC, a colon stricture, and the presence of postinflammatory pseudopolyps on colonoscopy.

■ CROHN'S DISEASE

Risk factors for developing cancer in Crohn's colitis are long-duration and extensive disease, bypassed colon segments, colon strictures, PSC, and family history of colon cancer. The cancer risks in CD and UC are probably equivalent for similar extent and duration of disease. In the CESAME study, a prospective observational cohort of IBD patients in France, the standardized incidence ratios of colorectal cancer were 2.2 for all IBD patients (95% CI, 1.5–3.0; $p < .001$) and 7.0 for patients with long-standing extensive colitis (both Crohn's and UC) (95% CI, 4.4–10.5; $p < .001$). Thus, the same endoscopic surveillance strategy used for UC is recommended for patients with chronic Crohn's colitis. A pediatric colonoscope can be used to pass narrow strictures in CD patients, but surgery should be considered in symptomatic patients with impassable strictures.

■ MANAGEMENT OF DYSPLASIA AND CANCER

Dysplasia can be flat or polypoid. If flat high-grade dysplasia is encountered on colonoscopic surveillance, the usual treatment is colectomy for UC and either colectomy or segmental resection for CD. If flat low-grade dysplasia is found (Fig. 326-13), most investigators recommend immediate colectomy. Adenomas may occur coincidentally in UC and CD patients with chronic colitis and can be removed endoscopically provided that biopsies of the surrounding mucosa are free of dysplasia.

IBD patients are also at greater risk for other malignancies. Patients with CD may have an increased risk of NHL, leukemia, and myelodysplastic syndromes. Severe, chronic, complicated perianal disease in CD patients may be associated with an increased risk of cancer in the lower rectum and anal canal (squamous cell cancers). Although the absolute risk of small-bowel adenocarcinoma in CD is low (2.2% at 25 years in one study), patients with long-standing, extensive, small-bowel disease should be considered for screening.

COVID 19 AND IBD

COVID-19, caused by SARS-CoV-2, was first reported in December 2019 and has rapidly spread throughout the world, leading to an international pandemic. Glucocorticoids, immunomodulators (thiopurines, MTX), biologics, and JAK inhibitors, commonly used to treat IBD, are associated with higher rates of serious viral and bacterial infections, and patients with IBD using these medications are potentially at increased risk of a serious COVID-19 infection. Yet, it is also possible that some forms of immune suppression may blunt the excessive immune response/cytokine storm characteristic of severe COVID-19 infection and consequently reduce mortality. Using data from the Surveillance



FIGURE 326-13 Medium-power view of low-grade dysplasia in a patient with chronic ulcerative colitis. Low-grade dysplastic crypts are interspersed among regenerating crypts. (Courtesy of Dr. R. Odze, Division of Gastrointestinal Pathology, Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts; with permission.)

2490 Epidemiology of Coronavirus Under Research Exclusion for Inflammatory Bowel Disease, it was found that increasing age (adjusted OR 1.04; 95% CI, 1.01–1.02), two or more comorbidities (adjusted OR 2.9; 95% CI, 1.1–7.8), and systemic glucocorticoids (adjusted OR 6.9; 95% CI, 2.3–20.5) are associated with severe COVID-19 in IBD patients. Anti-TNF treatment was not associated with severe COVID-19 (adjusted OR 0.9; 95% CI, 0.4–2.2).

FURTHER READING

- A B et al: High-definition chromoendoscopy superior to high-definition white-light endoscopy in surveillance of inflammatory bowel diseases in a randomized trial. *Clin Gastroenterol Hepatol* 18:2101, 2020.
- A A et al: Changing global epidemiology of inflammatory bowel disease: Sustaining health care delivery into the 21st century. *Clin Gastroenterol Hepatol*. 2020;18(6):1252–1260.
- B E, H ST: Checkpoint-inhibitor-induced colitis. *Am J Gastroenterol* 115:202, 2020.
- B CM et al: Events within the first year of life, but not the neonatal period, affect risk for later development of inflammatory bowel diseases. *Gastroenterology* 156:2190, 2019.
- B E et al: Corticosteroids, but not TNF antagonists, are associated with adverse COVID-19 outcomes in patients with inflammatory bowel diseases: Results from an international registry. *Gastroenterology* 159:481, 2020.
- G DB, X RJ: Pathway paradigms revealed from the genetics of inflammatory bowel disease. *Nature* 578:527, 2020.
- L A et al: Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. *Gastroenterology* 157:440, 2019.
- M U et al: Pregnancy and neonatal outcomes after fetal exposure to biologics and thiopurines among women with inflammatory bowel disease. *Gastroenterology* 160:1131, 2021.
- M FT et al: Familial risk of inflammatory bowel disease: A population-based cohort study 1977–2011. *Am J Gastroenterol* 110:564, 2015.
- N SC et al: Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: A systematic review of population-based studies. *Lancet* 390:2769, 2018.
- S BE et al: UNIFI Study Group. Ustekinumab as induction and maintenance therapy for ulcerative colitis. *N Engl J Med* 381:1201, 2019.
- S BE et al: VARSITY Study Group. Vedolizumab versus adalimumab for moderate-to-severe ulcerative colitis. *N Engl J Med* 381:1215, 2019.
- S S et al: AGA technical review on the management of moderate to severe ulcerative colitis. *Gastroenterology* 158:1465, 2020.

327

Irritable Bowel Syndrome

Chung Owyang

TABLE 327-1 Rome IV Diagnostic Criteria for Irritable Bowel Syndrome^a

Recurrent abdominal pain, on average, at least 1 day per week in the last 3 months, associated with ≥2 of the following criteria:

1. Related to defecation
2. Associated with a change in frequency of stool
3. Associated with a change in form (appearance) of stool

^aCriteria fulfilled for the last 3 months with symptom onset at least 6 months prior to diagnosis.

and genitourinary symptoms. Severity of symptoms varies and can significantly impair quality of life, resulting in high health care costs. Altered gastrointestinal (GI) motility, visceral hyperalgesia, disturbance of brain–gut interaction, abnormal central processing, autonomic and hormonal events, genetic and environmental factors, and psychosocial disturbances are variably involved, depending on the individual. This progress may result in improved methods of treatment.

CLINICAL FEATURES

IBS is a disorder that affects all ages, although most patients have their first symptoms before age 45. Women are diagnosed with IBS two to three times as often as men and make up 80% of the population with severe IBS. As indicated in Table 327-1, pain is a key symptom for the diagnosis of IBS. This symptom should be associated with defecation and/or have its onset associated with a change in frequency or form of stool. In comparison to Rome III, the Rome IV criteria are more stringent, requiring abdominal pain to occur at a minimum of once a week and eliminating “discomfort” as one of the criteria. Painless diarrhea or constipation does not fulfill the diagnostic criteria to be classified as IBS. Supportive symptoms that are not part of the diagnostic criteria include defecation straining, urgency or a feeling of incomplete bowel movement, passing mucus, and bloating.

Abdominal Pain According to the current IBS diagnostic criteria, abdominal pain is a prerequisite clinical feature of IBS. Abdominal pain in IBS is highly variable in intensity and location. It is frequently episodic and crampy, but it may be superimposed on a background of constant ache. Pain may be mild enough to be ignored, or it may interfere with daily activities. Despite this, malnutrition due to inadequate caloric intake is exceedingly rare with IBS. Sleep deprivation is also unusual because abdominal pain is almost uniformly present only during waking hours. Pain is often exacerbated by eating or emotional stress and improved by passage of flatus or stools. In addition, female patients with IBS commonly report worsening symptoms during the premenstrual and menstrual phases.

Altered Bowel Habits Alteration in bowel habits is the most consistent clinical feature in IBS. The most common pattern is constipation alternating with diarrhea, usually with one of these symptoms predominating. At first, constipation may be episodic, but eventually, it becomes continuous and increasingly intractable to treatment with laxatives. Stools are usually hard with narrowed caliber, possibly reflecting excessive dehydration caused by prolonged colonic retention and spasm. Most patients also experience a sense of incomplete evacuation, thus leading to repeated attempts at defecation in a short time span. Patients whose predominant symptom is constipation may have weeks or months of constipation interrupted with brief periods of diarrhea. In other patients, diarrhea may be the predominant symptom. Diarrhea resulting from IBS usually consists of small volumes of loose stools. Most patients have stool volumes of <200 mL. Nocturnal diarrhea does not occur in IBS. Diarrhea may be aggravated by emotional stress or eating. Stool may be accompanied by passage of large amounts of mucus. Bleeding is not a feature of IBS unless hemorrhoids are present, and malabsorption or weight loss does not occur.

Bowel pattern subtypes are highly unstable. In a patient population with ~33% prevalence rates of IBS-diarrhea predominant (IBS-D), IBS-constipation predominant (IBS-C), and IBS-mixed (IBS-M) forms, 75% of patients change subtypes, and 29% switch between IBS-C and IBS-D over 1 year.

Irritable bowel syndrome (IBS) is a functional bowel disorder characterized by abdominal pain or discomfort and altered bowel habits in the absence of detectable structural abnormalities. No clear diagnostic markers exist for IBS; thus, the diagnosis of the disorder is based on clinical presentation. In 2016, the Rome III criteria for the diagnosis of IBS were updated to Rome IV (Table 327-1). Throughout the world, ~10–20% of adults and adolescents have symptoms consistent with IBS. IBS symptoms tend to come and go over time and often overlap with other functional disorders such as fibromyalgia, headache, backache,

Gas and Flatulence Patients with IBS frequently complain of abdominal distention and increased belching or flatulence, all of which they attribute to increased gas. Although some patients with these symptoms actually may have a larger amount of gas, quantitative measurements reveal that most patients who complain of increased gas generate no more than a normal amount of intestinal gas. Most IBS patients have impaired transit and tolerance of intestinal gas loads. In addition, patients with IBS tend to reflux gas from the distal to the more proximal intestine, which may explain the belching. Some patients with bloating may also experience visible distention with increase in abdominal girth.

Upper GI Symptoms Between 25 and 50% of patients with IBS complain of dyspepsia, heartburn, nausea, and vomiting. This suggests that other areas of the gut apart from the colon may be involved. Prolonged ambulant recordings of small-bowel motility in patients with IBS show a high incidence of abnormalities in the small bowel during the diurnal (waking) period; nocturnal motor patterns are not different from those of healthy controls. The overlap between dyspepsia and IBS is great. The prevalence of IBS is higher among patients with dyspepsia (31.7%) than among those who reported no symptoms of dyspepsia (7.9%). Conversely, among patients with IBS, 55.6% reported symptoms of dyspepsia. In addition, the functional abdominal symptoms can change over time. Those with predominant dyspepsia or IBS can flux between the two. Thus, it is conceivable that functional dyspepsia and IBS are two manifestations of a single, more extensive digestive system disorder. Furthermore, IBS symptoms are prevalent in noncardiac chest pain patients, suggesting overlap with other functional gut disorders.

■ PATHOPHYSIOLOGY

The pathogenesis of IBS is poorly understood, although roles of abnormal gut motor and sensory activity, central neural dysfunction, psychological disturbances, mucosal inflammation, stress, and luminal factors such as bile acid malabsorption and gut dysbiosis have been proposed (Fig. 327-1).

GI Motor Abnormalities Studies of colonic myoelectrical and motor activity under unstimulated conditions have not shown consistent abnormalities in IBS. In contrast, colonic motor abnormalities are more prominent under stimulated conditions in IBS. IBS patients may exhibit increased rectosigmoid motor activity for up to 3 h after eating. Similarly, inflation of rectal balloons both in IBS-D and IBS-C patients leads to marked and prolonged distention-evoked contractile activity. Recordings from the transverse, descending, and sigmoid colon showed that the motility index and peak amplitude of high-amplitude propagating contractions (HAPCs) in diarrhea-prone IBS patients were greatly increased compared to those in healthy subjects and were

associated with rapid colonic transit and accompanied by abdominal pain.

Visceral Hypersensitivity IBS patients frequently exhibit exaggerated sensory responses to visceral stimulation. The frequency of perceptions of food intolerance is at least twofold more common than in the general population. Postprandial pain has been temporally related to entry of the food bolus into the cecum in 74% of patients. On the other hand, prolonged fasting in IBS patients is often associated with significant improvement in symptoms. Rectal balloon inflation produces nonpainful and painful sensations at lower volumes in IBS patients than in healthy controls without altering rectal tension, suggestive of visceral afferent dysfunction in IBS. Similar studies show gastric and esophageal hypersensitivity in patients with nonulcer dyspepsia and noncardiac chest pain, raising the possibility that these conditions have a similar pathophysiologic basis. Lipids lower the thresholds for the first sensation of gas, discomfort, and pain in IBS patients. Hence, postprandial symptoms in IBS patients may be explained in part by a nutrient-dependent exaggerated sensory component of the gastrocolonic response. In contrast to enhanced gut sensitivity, IBS patients do not exhibit heightened sensitivity elsewhere in the body. Thus, the afferent pathway disturbances in IBS appear to be selective for visceral innervation with sparing of somatic pathways. The mechanisms responsible for visceral hypersensitivity are still under investigation.

Central Neural Dysregulation The role of central nervous system (CNS) factors in the pathogenesis of IBS is strongly suggested by the clinical association of emotional disorders and stress with symptom exacerbation and the therapeutic response to therapies that act on cerebral cortical sites. Functional brain imaging studies such as magnetic resonance imaging (MRI) have shown that in response to distal colonic stimulation, the mid-cingulate cortex—a brain region concerned with attention processes and response selection—shows greater activation in IBS patients. Modulation of this region is associated with changes in the subjective unpleasantness of pain. In addition, IBS patients also show preferential activation of the prefrontal lobe, which contains a vigilance network within the brain that increases alertness. These may represent a form of cerebral dysfunction, leading to the increased perception of visceral pain.

Abnormal Psychological Features Abnormal psychiatric features are recorded in up to 80% of IBS patients, especially in referral centers; however, no single psychiatric diagnosis predominates. Most of these patients demonstrated exaggerated symptoms in response to visceral distention, and this abnormality persists even after exclusion of psychological factors.

Psychological factors influence pain thresholds in IBS patients, as stress alters sensory thresholds. An association between prior sexual or physical abuse and development of IBS has been reported. Clinical studies suggest that IBS has a strong developmental component that involves interactions of genetic and epigenetic factors early in life. These may modulate brain networks related to emotional arousal and/or central autonomic control, salience, and somatosensory integration. Abuse is associated with greater pain reporting, psychological distress, and poor health outcome. Brain functional MRI studies show greater activation of the posterior and middle dorsal cingulate cortex, which is implicated in affect processing in IBS patients with a past history of sexual abuse.

Postinfectious IBS A GI infection may predispose a patient to IBS. In an investigation of 544 patients with confirmed bacterial gastroenteritis, one-quarter developed IBS subsequently. Conversely, about a third of IBS patients experienced an acute “gastroenteritis-like” illness at the onset of their chronic IBS symptomatology. This group of “postinfective” IBS occurs more commonly in females and affects younger rather than older patients. Risk factors for developing postinfectious IBS include, in

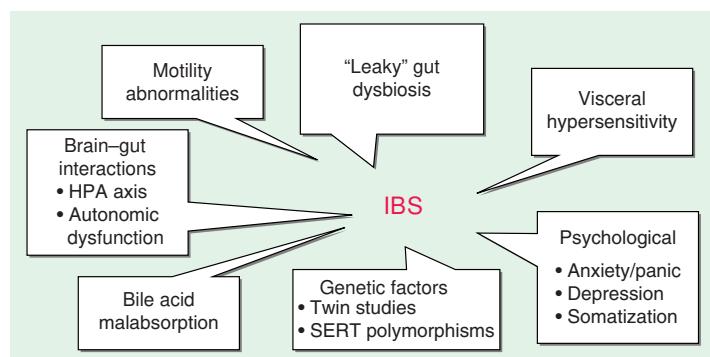


FIGURE 327-1 Pathophysiology of irritable bowel syndrome (IBS). The cause of IBS is likely to be multifactorial. Patients often show evidence of visceral hypersensitivity and motility abnormalities. Many IBS patients have increased anxiety and/or depression, and their symptoms are often exacerbated by mental or physical stress, suggesting abnormal brain-gut interaction. Genetic studies suggest a few IBS patients may have genetic abnormalities affecting the serotonin transport system in the enteric nerves. Up to 30% of IBS patients may have bile acid malabsorption. Gut dysbiosis and impaired mucosa permeability also have been reported in many IBS patients. This may lead to subclinical mucosa inflammation.

2492 order of importance, prolonged duration of initial illness, toxicity of infecting bacterial strain, smoking, mucosal markers of inflammation, female sex, depression, hypochondriasis, and adverse life events in the preceding 3 months. Age older than 60 years might protect against postinfectious IBS, whereas treatment with antibiotics has been associated with increased risk. The microbes involved in the initial infection are *Campylobacter*, *Salmonella*, and *Shigella*. Increased rectal mucosal enteroendocrine cells, T lymphocytes, and gut permeability are acute changes following *Campylobacter* enteritis that could persist for more than a year and may contribute to postinfective IBS.

Immune Activation and Mucosal Inflammation Some patients with IBS display persistent signs of low-grade mucosal inflammation with activated lymphocytes, mast cells, and enhanced expression of proinflammatory cytokines. Other studies also indicate that peripheral blood mononuclear cells (PBMCs) from IBS patients show abnormal release of proinflammatory cytokines such as interleukin (IL) 6, IL-1 β , and tumor necrosis factor (TNF). These abnormalities may contribute to abnormal epithelial secretion and visceral hypersensitivity. Located at the host–environment interface, mast cells are in close proximity to sensory nerves. Electromicroscopic evidence of mast cell activation is commonly observed in the colonic mucosa of IBS patients. Recent studies show that the proximity of activated mast cells to submucosal nerve fibers correlates with the frequency and severity of abdominal pain in patients with IBS. Other studies report that the colonic mucosa of IBS patients releases increased amounts of mast cell mediators, including histamine, proteases, and prostaglandin E₂. These findings, together with the observation that marked excitation of visceral sensory nerves innervating the colon occurs after exposure to IBS mucosal supernatant, support a prominent role for mast cells in the pathogenesis of visceral hypersensitivity. Increasing evidence suggests that some members of the superfamily of transient receptor potential (TRP) cation channels such as TRPV1 (vanilloid) channels are central to the initiation and persistence of visceral hypersensitivity. Mucosal inflammation can lead to increased expression of TRPV1 in the enteric nervous system. Enhanced expression of TRPV1 channels in the sensory neurons of the gut has been observed in IBS, and such expression appears to correlate with visceral hypersensitivity and abdominal pain. Interestingly, clinical studies have also shown increased intestinal permeability in patients with IBS-D. Psychological stress and anxiety can increase the release of proinflammatory cytokine, and this in turn may alter intestinal permeability. A clinical study showed that 39% of IBS-D patients had increased intestinal permeability as measured by the lactulose/mannitol ratio. These IBS patients also demonstrated a higher Functional Bowel Disorder Severity Index (FBDSI) score and increased hypersensitivity to visceral nociceptive pain stimuli. This provides a functional link among psychological stress, immune activation, and symptom generation in patients with IBS.

Altered Gut Flora A high prevalence of small-intestinal bacterial overgrowth in IBS patients has been noted based on positive lactulose hydrogen breath test. This finding, however, has been challenged by a number of other studies that found no increased incidence of bacterial overgrowth based on jejunal aspirate culture. Abnormal H₂ breath test can occur because of small-bowel rapid transit and may lead to erroneous interpretation. Hence, the role of testing for small-intestinal bacterial overgrowth in IBS patients remains unclear.

Studies using culture-independent approaches such as 16S rRNA gene-based analysis found significant differences between the molecular profile of the fecal microbiota of IBS patients and that of healthy subjects. A review of 24 studies involving 827 IBS patients showed extensive variability in bacterial flora among IBS patients and healthy subjects. However, a few general observations were made: (1) an increase in the ratio of fecal Firmicutes to fecal Bacteroidetes was noted in IBS; (2) the diversity of the microbiota was decreased; and (3) these changes were accompanied by an increase in instability of the bacteria flora in IBS. Despite a lack of consensus on the exact microbial difference between IBS patients and controls, IBS patients generally had decreased proportions of the genera *Bifidobacterium* and *Faecalibacterium* and increased abundance of family Enterobacteriaceae

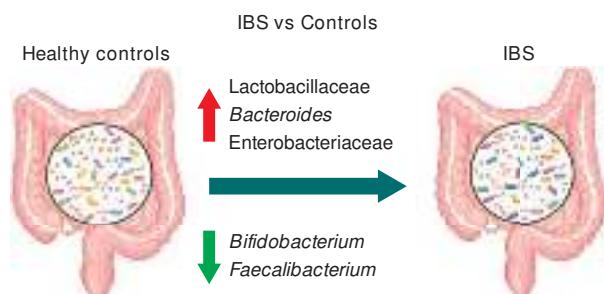


FIGURE 327-2 Changes in gut microbiota among patients with irritable bowel syndrome. (Adapted from R Pittayanon et al: *Gastroenterology* 157:97, 2019.)

(phylum Proteobacteria), family Lactobacillaceae, and genus *Bacteroides* (phylum Bacteroidetes) (Fig. 327-2). Many members of the genus *Faecalibacterium* are butyrate-producing and anti-inflammatory organisms and may reduce IBS symptoms via mediation of IL-17 expression. Similarly, members of the genus *Bifidobacterium* are also anti-inflammatory organisms and may reduce mucosal inflammation in IBS patients in clinical trials. On the other hand, the three groups of bacteria that were increased were potentially harmful commensal microbiota. The gram-negative Enterobacteriaceae family is capable of injuring epithelium lining and inducing mucosal inflammation via a lipopolysaccharide-dependent pathway. Members of the genus *Bacteroides* such as *Bacteroides fragilis* produces toxin to dissolve glycoproteins and induce mucosal inflammation. Lastly, in the family of Lactobacillaceae, *Lactobacillus* can produce gas and organic acids from glucose and fructose fermentation, resulting in bloating and abdominal pain. It is conceivable that gut dysbiosis acting in concert with genetic susceptibility and environmental insults may alter mucosal permeability and increase antigen presentation to the immune cells in the lamina propria. This may result in mast cell activation and altered enteric neuronal and smooth-muscle function causing IBS symptoms. In addition, release of cytokines and chemokines from mucosal inflammation may generate extra-GI symptoms such as chronic fatigue, muscle pain, and anxiety (Fig. 327-3).

Abnormal Serotonin Pathways The serotonin-containing enterochromaffin cells in the colon are increased in a subset of IBS-D patients compared to healthy individuals or patients with ulcerative colitis. Furthermore, postprandial plasma serotonin levels were significantly higher in this group of patients compared to healthy controls. Tryptophan hydroxylase 1 (TPH1) is the rate-limiting enzyme in enterochromaffin cell serotonin biosynthesis, and functional TPH1 polymorphism has been shown to be associated with IBS habit subtypes. In addition, gut microbes promote colonic serotonin production through an effect of short-chain fatty acids on enterochromaffin cells. In IBS patients, the expression of mucosal serotonin reuptake transporter (SERT) is downregulated due to gram negative gut dysbiosis. Thus, gut and reuptake dysbiosis in IBS may contribute to abnormal serotonin synthesis in this disorder. Because serotonin plays an important role in the regulation of GI motility and visceral perception, the increased release of serotonin may contribute to the postprandial symptoms of these patients and provides a rationale for the use of serotonin antagonists in the treatment of this disorder.

APPROACH TO THE PATIENT

Irritable Bowel Syndrome

Because IBS is a disorder for which no pathognomonic abnormalities have been identified, its diagnosis relies on recognition of positive clinical features and elimination of other organic diseases. Symptom-based criteria have been developed for the purpose of differentiating patients with IBS from those with organic diseases. These include the Manning, Rome I, Rome II, Rome III, and Rome IV criteria. Rome IV criteria for the diagnosis of IBS were published in 2016 (Table 327-1) and defined IBS on the basis of abdominal

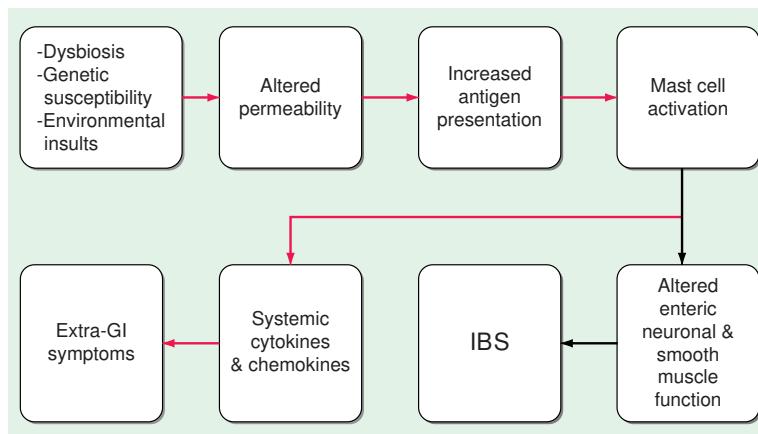


FIGURE 327-3 Gut dysbiosis and irritable bowel syndrome (IBS). Gut dysbiosis acting in concert with genetic and environmental factors may alter intestinal permeability and increase antigen presentation, resulting in mast cell activation. Products of mast cell degranulation may alter neuronal and smooth-muscle function causing IBS symptoms. The cytokines and chemokines generated from mucosal inflammation may cause symptoms such as fibromyalgia, chronic fatigue, and mood changes. GI, gastrointestinal. (Adapted from NJ Talley, AA Fodor: *Gastroenterology* 141:1555, 2011.)

pain and altered bowel habits that occur with sufficient frequency in affected patients. A careful history and physical examination are frequently helpful in establishing the diagnosis. Clinical features suggestive of IBS include recurrence of lower abdominal pain with altered bowel habits over a period of time without progressive deterioration, onset of symptoms during periods of stress or emotional upset, absence of other systemic symptoms such as fever and weight loss, and small-volume stool without any evidence of blood.

On the other hand, the appearance of the disorder for the first time in old age, progressive course from time of onset, persistent diarrhea after a 48-h fast, and presence of nocturnal diarrhea or steatorrheal stools argue against the diagnosis of IBS.

Because the major symptoms of IBS—abdominal pain, abdominal bloating, and alteration in bowel habits—are common complaints of many GI organic disorders, the list of differential diagnoses is a long one. The quality, location, and timing of pain may be helpful to suggest specific disorders. Pain due to IBS that occurs in the epigastric or perumbilical area must be differentiated from biliary tract disease, peptic ulcer disorders, intestinal ischemia, and carcinoma of the stomach and pancreas. If pain occurs mainly in the lower abdomen, the possibility of diverticular disease of the colon, inflammatory bowel disease (including ulcerative colitis and Crohn's disease), and carcinoma of the colon must be considered. Postprandial pain accompanied by bloating, nausea, and vomiting suggests gastroparesis or partial intestinal obstruction. Patients with small intestinal bacteria overgrowth can present with abdominal pain, nausea, and bloating, and this possibility should be ruled out before making a diagnosis of IBS. Intestinal infestation with *Giardia lamblia* or other parasites may cause similar symptoms. When diarrhea is the major complaint, the possibility of lactase deficiency, laxative abuse, malabsorption, celiac sprue, hyperthyroidism, inflammatory bowel disease, and infectious diarrhea must be ruled out. On the other hand, constipation may be a side effect of many different drugs, such as anticholinergic, antihypertensive, and antidepressant medications. Endocrinopathies such as hypothyroidism and hypoparathyroidism must also be considered in the differential diagnosis of constipation. In addition, acute intermittent porphyria and lead poisoning may present in a fashion similar to IBS, with painful constipation as the major complaint. These possibilities are suspected on the basis of their clinical presentations and are confirmed by appropriate serum and urine tests.

Few tests are required for patients who have typical IBS symptoms and no alarm features. Unnecessary investigations may be costly and even harmful. The American Gastroenterological Association has delineated factors to be considered when determining

the aggressiveness of the diagnostic evaluation. These include the duration of symptoms, the change in symptoms over time, the age and sex of the patient, the referral status of the patient, prior diagnostic studies, a family history of colorectal malignancy, and the degree of psychosocial dysfunction. Thus, a younger individual with mild symptoms requires a minimal diagnostic evaluation, while an older person or an individual with rapidly progressive symptoms should undergo a more thorough exclusion of organic disease. Most patients should have a complete blood count and sigmoidoscopic examination; in addition, stool specimens should be examined for ova and parasites in those who have diarrhea. In patients with persistent diarrhea not responding to simple anti-diarrheal agents, a sigmoid colon biopsy should be performed to rule out microscopic colitis. In those age >40 years, an air-contrast barium enema or colonoscopy should also be performed. If the main symptoms are diarrhea and increased gas, the possibility of lactase deficiency should be ruled out with a hydrogen breath test or with evaluation after a 3-week lactose-free diet. Excessive gas with bloating also raises the possibility of small-bowel bacteria overgrowth and should be ruled out with a glucose hydrogen breath test. Some patients with IBS-D may have undiagnosed celiac sprue. Because the symptoms of celiac sprue respond to a gluten-free diet, testing for celiac sprue in IBS may prevent years of morbidity and attendant expense. Decision-analysis studies show that serology testing for celiac sprue in patients with IBS-D has an acceptable cost when the prevalence of celiac sprue is >1% and is the dominant strategy when the prevalence is >8%. In patients with concurrent symptoms of dyspepsia, upper GI radiographs or esophagogastroduodenoscopy may be advisable. In patients with postprandial right upper quadrant pain, an ultrasonogram of the gallbladder should be obtained. Laboratory features that argue against IBS include evidence of anemia, elevated sedimentation rate, presence of leukocytes or blood in stool, and stool volume >200–300 mL/d. These findings would necessitate other diagnostic considerations.

TREATMENT

Irritable Bowel Syndrome

PATIENT COUNSELING AND DIETARY ALTERATIONS

Reassurance and careful explanation of the functional nature of the disorder and of how to avoid obvious food precipitants are important first steps in patient counseling and dietary change. Occasionally, a meticulous dietary history may reveal substances (such as coffee, disaccharides, legumes, and cabbage) that aggravate symptoms. Excessive fructose and artificial sweeteners, such as

TABLE 327-2 Some Common Food Sources of FODMAPs

FOOD TYPE	FREE FRUCTOSE	LACTOSE	FRUCTANS	GALACTO-Oligosaccharides	POLYOLS
Fruits	Apple, cherry, mango, pear, watermelon		Peach, persimmon, watermelon		Apple, apricot, pear, avocado, blackberries, cherry, nectarine, plum, prune
Vegetables	Asparagus, artichokes, sugar snap peas		Artichokes, beetroot, Brussels sprout, chicory, fennel, garlic, leek, onion, peas		Cauliflower, mushroom, snow peas
Grains and cereals			Wheat, rye, barley		
Nuts and seeds			Pistachios		
Milk and milk products		Milk, yogurt, ice cream, custard, soft cheeses			
Legumes			Legumes, lentils, chickpeas	Legumes, chickpeas, lentils	
Other	Honey, high-fructose corn syrup		Chicory drinks		
Food additives			Inulin, FOS		Sorbitol, mannitol, maltitol, xylitol, isomalt

Abbreviations: FODMAPs, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; FOS, fructo-oligosaccharides.

Source: Reproduced with permission from PR Gibson et al: Food choice as a key management strategy for functional gastrointestinal symptoms. Am J Gastroenterol 107:657, 2012.

sorbitol or mannitol, may cause diarrhea, bloating, cramping, or flatulence. As a therapeutic trial, patients should be encouraged to eliminate any foodstuffs that appear to produce symptoms. However, patients should avoid nutritionally depleted diets. A diet low in fermentable oligosaccharides, disaccharides, monosaccharides, and polyols (FODMAPs) (Table 327-2) has been shown to be helpful in IBS patients (see “Low FODMAP Diet”).

STOOL-BULKING AGENTS

High-fiber diets and bulking agents, such as bran or hydrophilic colloid, are frequently used in treating IBS. The water-holding action of fibers may contribute to increased stool bulk because of the ability of fiber to increase fecal output of bacteria. Fiber also speeds up colonic transit in most people. In diarrhea-prone patients, whole-colonic transit is faster than average; however, dietary fiber can also delay transit. Furthermore, because of their hydrophilic properties, stool-bulking agents bind water and thus prevent both excessive hydration and dehydration of stool. The latter observation may explain the clinical experience that a high-fiber diet relieves diarrhea in some IBS patients. Fiber supplementation with psyllium has been shown to reduce perception of rectal distention, indicating that fiber may have a positive effect on visceral afferent function.

Controlled trials of dietary fiber in IBS patients have produced variable results. This is not surprising since IBS is a heterogeneous disorder, with some patients being constipated and others having predominant diarrhea. Most investigations report increases in stool weight, decreases in colonic transit times, and improvement in constipation. Others have noted benefits in patients with alternating diarrhea and constipation, pain, and bloating. However, most studies observe no response in patients with diarrhea- or pain-predominant IBS. Compared to insoluble dietary fiber such as wheat bran, soluble fibers such as psyllium preparations tend to produce less bloating and distention. Fiber should be started at a nominal dose and slowly titrated up as tolerated over the course of several weeks to a targeted dose of 20–30 g of total dietary and supplementary fiber per day. Even when used judiciously, fiber can exacerbate bloating, flatulence, constipation, and diarrhea. Patients with drug-induced or slow colonic transit constipation usually do not respond to fiber supplementation.

ANTISPASMODICS

Clinicians have observed that anticholinergic drugs may provide temporary relief for symptoms such as painful cramps related to intestinal spasm. Although controlled clinical trials have produced mixed results, evidence generally supports the beneficial effects

of anticholinergic drugs for pain. Physiologic studies demonstrate that anticholinergic drugs inhibit the gastrocolic reflex; hence, postprandial pain is best managed by giving antispasmodics 30 min before meals so that effective blood levels are achieved shortly before the anticipated onset of pain. Most anticholinergics contain natural belladonna alkaloids, which may cause xerostomia, urinary hesitancy and retention, blurred vision, and drowsiness. They should be used in the elderly with caution. Some physicians prefer to use synthetic anticholinergics such as dicyclomine that have less effect on mucous membrane secretion and produce fewer undesirable side effects. Peppermint oil appears to reduce abdominal cramps by some undefined mechanism. In a meta-analysis of nine double-blind randomized controlled trials evaluating 726 IBS patients, peppermint oil was found to be significantly superior to placebo for global improvement of IBS symptoms and reduction in abdominal pain. The most commonly reported adverse event was heartburn, which was mild and transient.

ANTIDIARRHEAL AGENTS

Peripherally acting opiate-based agents are the initial therapy of choice for IBS-D. Physiologic studies demonstrate increases in segmenting colonic contractions, delays in fecal transit, increases in anal pressures, and reductions in rectal perception with these drugs. When diarrhea is severe, especially in the painless diarrhea variant of IBS, small doses of loperamide, 2–4 mg every 4–6 h up to a maximum of 12 mg/d, can be prescribed. These agents are less addictive than paretic, codeine, or tincture of opium. In general, the intestines do not become tolerant of the antidiarrheal effect of opiates, and increasing doses are not required to maintain antidiarrheal potency. These agents are most useful if taken before anticipated stressful events that are known to cause diarrhea. However, not infrequently, a high dose of loperamide may cause cramping because of increases in segmenting colonic contractions. Another antidiarrheal agent that may be used in IBS patients is the bile acid binder cholestyramine resin because up to 30% of IBS-D patients may have bile acid malabsorption.

ANTIDEPRESSANT DRUGS

In addition to their mood-elevating effects, antidepressant medications have several physiologic effects that suggest they may be beneficial in IBS. In IBS-D patients, the tricyclic antidepressant imipramine slows jejunal migrating motor complex transit propagation and delays orocecal and whole-gut transit, indicative of a motor inhibitory effect. Some studies also suggest that tricyclic agents may alter visceral afferent neural function.

A number of studies indicate that tricyclic antidepressants may be effective in some IBS patients. When stratified according to the predominant symptoms, improvements were observed in IBS-D patients, with no improvement being noted in IBS-C patients. The beneficial effects of the tricyclic compounds in the treatment of IBS appear to be independent of their effects on depression. The therapeutic benefits for the bowel symptoms occur faster and at a lower dosage. The efficacy of antidepressant agents from other chemical classes in the management of IBS is less well evaluated. In contrast to tricyclic agents, the selective serotonin reuptake inhibitor (SSRI) paroxetine accelerates orocecal transit, raising the possibility that this drug class may be useful in IBS-C patients. The SSRI citalopram blunts perception of rectal distention and reduces the magnitude of the gastrocolonic response in healthy volunteers. A small placebo-controlled study of citalopram in IBS patients reported reductions in pain. However, these findings could not be confirmed in another randomized controlled trial. Hence, the efficacy of SSRIs in the treatment of IBS needs further confirmation.

ANTIFLATULENCE THERAPY

The management of excessive gas is seldom satisfactory, except when there is obvious aerophagia or disaccharidase deficiency. Patients should be advised to eat slowly and not chew gum or drink carbonated beverages. Bloating may decrease if an associated gut syndrome such as IBS or constipation is improved. If bloating is accompanied by diarrhea and worsens after ingesting dairy products, fresh fruits, vegetables, or juices, further investigation or a dietary exclusion trial may be worthwhile. Avoiding flatogenic foods, exercising, losing excess weight, and taking activated charcoal are safe but unproven remedies. A low FODMAP diet has been shown to be quite effective to reduce gas and bloating (see “Low FODMAP Diet”). Data regarding the use of surfactants such as simethicone are conflicting. Antibiotics may help in a subgroup of IBS patients with predominant symptoms of bloating. Beano, an over-the-counter oral β -glycosidase solution, may reduce rectal passage of gas without decreasing bloating and pain. Pancreatic enzymes reduce bloating, gas, and fullness during and after high-calorie, high-fat meal ingestion.

SEROTONIN RECEPTOR MODULATORS

Serotonin 5-HT₃ and 5-HT₄ receptors are found throughout the GI tract. Prucalopride, a dihydrobenzo-furancarboxamide derivative, is a new selective agonist of 5-HT₄. In six of seven multicenter, double-blind, randomized trials of prucalopride in patients with chronic constipation, the drug was more effective than placebo. The most frequently encountered side effects were headache, nausea, and diarrhea, which were mostly transient. Unlike with the older 5-HT₄ agonist tegaserod, there were no significant cardiovascular side effects. Prucalopride was approved by the European Medicines Agency and the U.S. Food and Drug Administration (FDA) for treatment of chronic constipation.

Another 5-HT₄ receptor agonist, tegaserod, also exhibits prokinetic activity by stimulating peristalsis. Clinical studies involving >4000 IBS-C patients reported reduction in abdominal discomfort and improvements in constipation and bloating compared to placebo. Diarrhea is the only major side effect. In 2007, the drug was voluntarily withdrawn from the market after a greater number of cardiovascular complications were observed in a database of 18,000 patients receiving tegaserod (0.11 vs 0.01% in placebo). In 2019, the FDA reviewed additional data and approved the use of tegaserod in women younger than 65 years old who do not have a history of ischemic cardiovascular disease and who have no more than one risk factor for cardiovascular disease.

SECRETAGOGUES

Lubiprostone, linaclotide, and plecanatide are secretagogues that stimulate net efflux of ions and water into the intestinal lumen and thus enhance transit and facilitate ease of defecation. By activating channels on the apical (luminal) enterocyte surface, these secretagogues increase intestinal chloride secretion. Other ion channels

and transporters secrete sodium into the intestine to maintain electroneutrality, followed by the secretion of water. Lubiprostone is a bicyclic fatty acid derived from prostaglandin E, that activates type 3 chloride channels in the apical membrane of intestinal epithelial cells. Oral lubiprostone was effective in the treatment of patients with IBS-C in large phase 3, randomized, double-blind, placebo-controlled multicenter trials. The recommended daily dose is 24 mg twice daily. In general, the drug is quite well tolerated. The major side effects are nausea and diarrhea. Linaclotide and plecanatide are minimally absorbed 14-amino-acid peptide guanylate cyclase-C (GC-C) agonists that bind to and activate GC-C on the luminal surface of intestinal epithelium. The subsequent increase in cyclic guanosine monophosphate activates the cyclic fibrosis transmembrane regulator and induces fluid secretion into the GI tract. These drugs are similar to endogenous peptides secreted by the small intestine (uroguanylin) or colon (guanylin). In two 12-week, double-blind, randomized, controlled trials, linaclotide (290 or 145 μ g, once daily) reduced constipation and pain. A lower dose (72 μ g once daily) was also more effective than placebo. Linaclotide has been approved by the FDA for treatment of constipation in IBS-C patients. Plecanatide (3- and 6-mg doses) also has been shown to be more effective than placebo in two phase 3 trials. The 3-mg once-daily dose has been approved by the FDA. The only significant side effect was diarrhea, which occurred in <5% of patients. Linaclotide and plecanatide are of similar efficacy and tolerability for the treatment of chronic constipation. Tenapanor, a small-molecule inhibitor of GI sodium-hydrogen exchange-3, has been shown to be more effective than placebo when given at 50 mg twice daily in patients with IBS-C.

OSMOTIC LAXATIVES

Osmotic agents such as magnesium citrate-based products, sodium phosphate-based products, and nonabsorbable carbohydrates are hypertonic products that, through osmosis, extract fluid into the intestinal lumen to soften stool and enhance colonic transit. In contrast, polyethylene glycol (PEG)-based solution is iso-osmotic and induces bowel movement by high-volume lavage. The osmotic laxatives were better than placebo in improving symptoms of chronic constipation in clinical trials. However, chronic use of magnesium hydroxide may result in severe hypermagnesemia in patients with renal impairment. Frequent sodium phosphate-based bowel cleansing should be avoided as this is associated with hyperphosphatasemia, hypocalcemia, and hypokalemia. In 19 trials, PEG consistently induced more bowel movements than placebo. A Cochrane review of 10 randomized trials showed that PEG was superior to lactulose for improving stool frequency and abdominal pain. Among the nonabsorbable carbohydrates, lactulose and sorbitol had similar laxative effects. However, bacterial metabolism of unabsorbed carbohydrates often leads to gas production and abdominal pain, which can limit long-term use.

MODULATION OF GUT FLORA

Because altered colonic flora (gut dysbiosis) may contribute to the pathogenesis of IBS, this has led to great interest in using antibiotics, prebiotics, probiotics, and dietary measures to treat IBS.

Antibiotics Antibiotic treatment benefits a subset of IBS patients. In a double-blind, randomized, placebo-controlled study, neomycin dosed at 500 mg twice daily for 10 days was more effective than placebo at improving symptom scores among IBS patients. The nonabsorbed oral antibiotic rifaximin is the most thoroughly studied antibiotic for the treatment of IBS. In a double-blind, placebo-controlled study, patients receiving rifaximin at a dose of 550 mg two times daily for 2 weeks experienced substantial improvement of global IBS symptoms over placebo. Rifaximin is the only antibiotic with demonstrated sustained benefit beyond therapy cessation in IBS patients. The drug has a favorable safety and tolerability profile compared with systemic antibiotics. A systematic review and meta-analysis of five studies of IBS patients found that rifaximin is more effective than placebo for global symptoms and bloating

(odds ratio 1.57), with a number needed to treat (NNT) of 10.2. The modest therapeutic gain was similar to that yielded by other currently available therapies for IBS. However, currently, there are still insufficient data to recommend routine use of this antibiotic in the treatment of IBS.

Prebiotics These are nondigestible food ingredients that stimulate growth and/or activity of bacteria in the GI tract. There have been four randomized trials to examine the effects of prebiotics. Three of the four studies reported that prebiotics worsened or did not improve IBS symptoms. This is not surprising given the adverse effects of a high-carbohydrate diet on IBS symptoms.

Probiotics These are defined as live microorganisms that when administered in adequate amounts confer a health benefit on the host. A meta-analysis of 10 probiotic studies in IBS patients found significant relief of pain and bloating with the use of *Bifidobacterium breve*, *Bifidobacterium longum*, and *Lactobacillus acidophilus* species compared to placebo. However, there was no change in stool frequency or consistency. Large-scale studies of well-phenotyped IBS patients are needed to establish the efficacy of these probiotics.

Low FODMAP Diet A diet rich in FODMAP (fermentable oligosaccharides, disaccharides, monosaccharides, and polyols) often triggers symptoms in IBS patients. FODMAPs are poorly absorbed by the small intestine and fermented by bacteria in the colon to produce gas and osmotically active carbohydrates (Fig. 327-4). At the same time, on entering the colon, FODMAPs may serve as nutrients for the colonic bacteria and promote the growth of gram-negative commensal bacteria, which may induce epithelial damage and subclinical mucosa inflammation. Fructose and fructans induce IBS symptoms in a dose-dependent manner. In contrast, a low FODMAP diet reduces IBS symptoms. A systematic review and meta-analysis of seven studies of IBS patients found that a low FODMAP diet was associated with reduced global symptoms compared with control interventions. These studies showed symptomatic benefit of restricting FODMAPs in 50–80% of patients with IBS. There is increasing support for recommending a low FODMAP diet as first-line treatment for IBS patients. Given that between 20 and 50% of patients do not respond to a low FODMAP

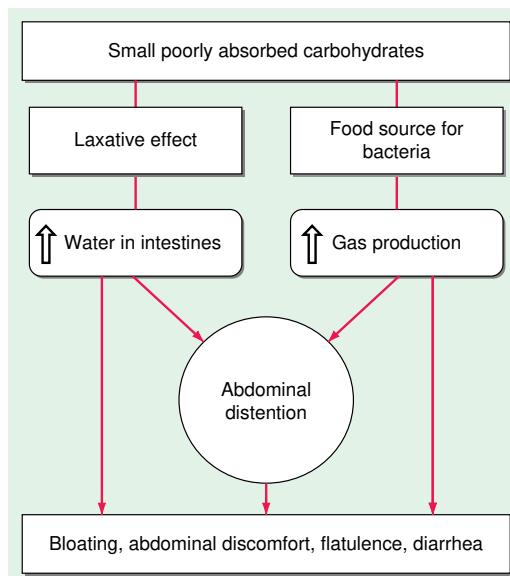


FIGURE 327-4 Pathogenesis of FODMAP-related symptoms. FODMAPs are poorly absorbed by the small intestine and fermented by gut bacteria to produce gas and osmotically active carbohydrates. These events act in concert to cause bloating, flatulence, and diarrhea. FODMAPs may also serve as nutrients for colonic bacteria, which may induce mucosa inflammation. FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols. (Figure created using data from <http://www.nutritiontoyou.com/wp-content/uploads/2014/06/IBS-symptoms.png>.)

TABLE 327-3 Spectrum of Severity in IBS

	MILD	MODERATE	SEVERE
Clinical Features			
Prevalence	70%	25%	5%
Correlations with gut physiology	+++	++	+
Symptoms constant	0	+	+++
Psychosocial difficulties	0	+	+++
Health care issues	+	++	+++
Practice type	Primary	Specialty	Referral

diet, the identification of patients who are more likely to respond to a low FODMAP diet at baseline would be highly beneficial. In a small clinical study comparing responders to a low FODMAP diet to nonresponders, fecal volatile organic acid profiling at baseline accurately predicted response in 97% of cases. This finding needs to be confirmed by large prospective cohort studies.

SUMMARY

The treatment strategy of IBS depends on the severity of the disorder (Table 327-3). Most IBS patients have mild symptoms. They are usually cared for in primary care practices, have little or no psychosocial difficulties, and do not seek health care often. Treatment usually involves education, reassurance, and dietary/lifestyle changes. A smaller portion have moderate symptoms that are usually intermittent and correlate with altered gut physiology, e.g., worsened with eating or stress and relieved by defecation. For IBS-D patients, treatments include gut-acting pharmacologic agents such as antispasmodics, antidiarrheals, bile acid binders, and the newer gut serotonin modulators (Table 327-4). In IBS-C patients, increased fiber intake and the use of osmotic agents such as polyethylene glycol may achieve satisfactory results. For patients with more severe constipation, a chloride channel opener (lubiprostone) or GC-C agonist (linaclotide or plecanatide) may be considered. For IBS patients with predominant gas and bloating, a low-FODMAP diet may provide significant relief. Some patients

TABLE 327-4 Possible Drugs for a Dominant Symptom in IBS

SYMPTOM	DRUG	DOSE
Diarrhea	Loperamide	2–4 mg when necessary/maximum 12 g/d
	Cholestyramine resin	4 g with meals
	Alosetron ^a	0.5–1 mg bid (for severe IBS, women)
Constipation	Psyllium husk	3–4 g bid with meals, then adjust
	Methylcellulose	2 g bid with meals, then adjust
	Calcium polycarbophil	1 g qd to qid
	Lactulose syrup	10–20 g bid
	70% sorbitol	15 mL bid
	Polyethylene glycol 3350	17 g in 250 mL water qd
	Lubiprostone (Amitiza)	24 mg bid
	Magnesium hydroxide	30–60 mL qd
	Linaclotide (Plecanatide)	290 µg qd/3 mg qd
Abdominal pain	Prucalopride	2 mg qd
	Smooth-muscle relaxant	qd to qid ac
	Tricyclic antidepressants	Start 25–50 mg hs, then adjust
Gas and bloating	Selective serotonin reuptake inhibitors	Begin small dose, increase as needed
	Low FODMAP diet	
	Probiotics	qd
	Rifaximin	550 mg bid

^aAvailable only in the United States.

Abbreviation: FODMAP, fermentable oligosaccharides, disaccharides, monosaccharides, and polyols; IBS, irritable bowel syndrome.

Source: Reproduced with permission from GF Longstreth et al: Functional bowel disorders. *Gastroenterology* 130:1480, 2006.

may benefit from probiotics and rifaximin treatment. A small proportion of IBS patients have severe and refractory symptoms, are usually seen in referral centers, and frequently have constant pain and psychosocial difficulties. This group of patients is best managed with antidepressants and other psychological treatments (Table 327-4). Clinical trials demonstrating success of a low FODMAP diet in improving IBS symptoms and quality of life provide strong evidence supporting the use of this dietary approach in the treatment of IBS. These observations, if confirmed, may lead to the use of the low FODMAP diet as the first line of treatment of IBS patients with moderate to severe symptoms.

FURTHER READING

- B AE, L BE: Mechanisms, evaluation, and management of chronic constipation. *Gastroenterology* 158:1232, 2020.
- D J et al: A systematic review and meta-analysis evaluating the efficacy of a gluten-free diet and a low FODMAP diet in treating symptoms of irritable bowel syndrome. *Am J Gastroenterol* 113:1290, 2018.
- D DA: Functional gastrointestinal disorders: History, pathophysiology, clinical features, and Rome IV. *Gastroenterology* 150:1262, 2016.
- M EA et al: Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology* 146:1500; 2014.
- P R et al: Gut microbiota in patients with irritable bowel syndrome: A systematic review. *Gastroenterology* 157:97, 2019.
- Z SY et al: FODMAP diet modulates visceral nociception by lipopolysaccharide-mediated intestinal barrier dysfunction and intestinal inflammation. *J Clin Invest* 128:267, 2018.

328

Diverticular Disease and Common Anorectal Disorders

Susan L. Gearhart

DIVERTICULAR DISEASE

Incidence and Epidemiology In the United States, diverticulosis affects one-third of the population aged >60 years, and in most instances, there are no associated symptoms. However, 10–25% of individuals with diverticulosis will develop acute diverticular disease. In addition, 10–25% of individuals with diverticular disease will experience recurrent symptoms, and up to 10% will develop complications leading to surgery. Diverticular disease has become the fifth most costly gastrointestinal disorder in the United States and is the leading indication for elective colon resection. The incidence of diverticular disease is on the rise, especially among individuals <40 years of age. The majority of patients with diverticular disease report a lower health-related quality of life and more depression as compared to matched controls, thus adding to health care costs. Formerly, diverticular disease was confined to developed countries; however, with the adoption of westernized diets in underdeveloped countries, diverticulosis is on the rise across the globe. Immigrants to the United States develop diverticular disease at the same rate as U.S. natives. Although the prevalence among females and males is similar, males tend to present at a younger age.

Anatomy and Pathophysiology Two types of diverticula occur in the intestine: true and false (or pseudodiverticula). A true diverticulum is a saclike herniation of the entire bowel wall, whereas a pseudodiverticulum involves only a protrusion of the mucosa and submucosa through the muscularis propria of the colon (Fig. 328-1). The type of

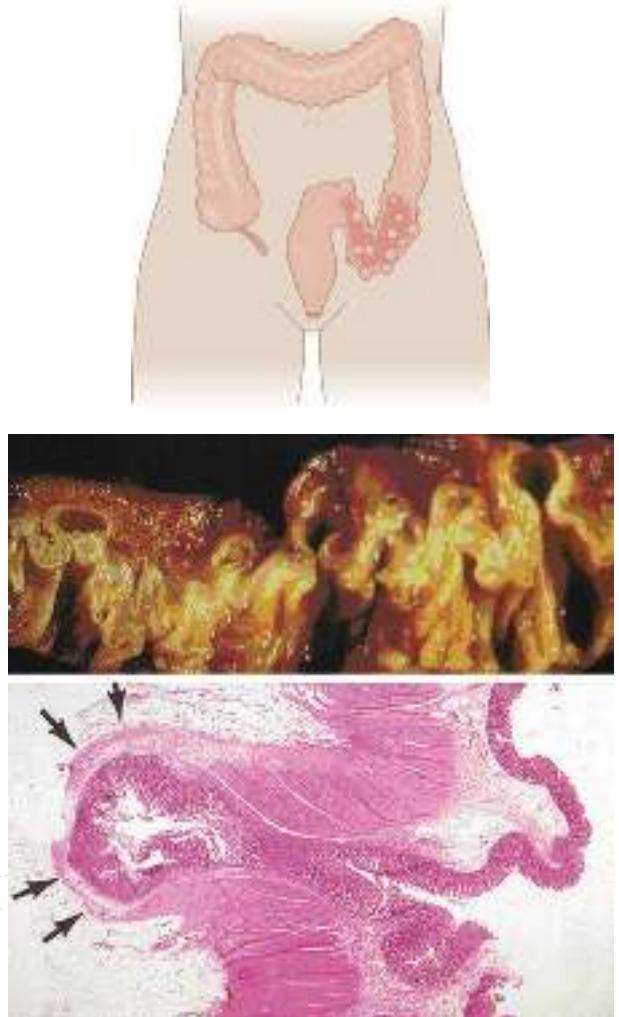


FIGURE 328-1 Gross and microscopic view of sigmoid diverticular disease. Arrows mark an inflamed diverticulum with the diverticular wall made up only of mucosa.

diverticulum most commonly affecting the colon is the pseudodiverticulum. The diverticula occur at the point where the nutrient artery, or vasa recta, penetrates through the muscularis propria, resulting in a break in the integrity of the colonic wall. This anatomic restriction may be a result of the relative high-pressure zone within the muscular sigmoid colon. Thus, higher-amplitude contractions combined with constipated, high-fat-content stool within the sigmoid lumen in an area of weakness in the colonic wall result in the creation of these diverticula. Consequently, the vasa recta is either compressed or eroded, leading to either perforation or bleeding.

Diverticula commonly affect the left and sigmoid colon; the rectum is always spared. However, in Asian populations, 70% of diverticula are seen in the right colon and cecum as well. Yamada et al. found right-sided colonic diverticulosis in 22% of Japanese patients undergoing colonoscopy. *Diverticulitis* is inflammation of a diverticulum. Previous understanding of the pathogenesis of diverticulosis attributed a low-fiber diet as the sole culprit, and onset of diverticulitis would occur acutely when these diverticula become obstructed. However, evidence now suggests that the pathogenesis is more complex and multifactorial. Better understanding of the gut microbiota suggests that dysbiosis is an important aspect of disease. Chronic low-grade inflammation is thought to play a key role in neuronal degeneration, leading to dysmotility and high intraluminal pressure. As a consequence, pockets or outpouchings develop in the colonic wall where it is weakest.

2498 Presentation, Evaluation, and Management of Diverticular Bleeding Hemorrhage from a colonic diverticulum is the most common cause of hematochezia in patients >60 years, yet only 20% of patients with diverticulosis will have gastrointestinal bleeding. Patients at increased risk for bleeding tend to be hypertensive, have atherosclerosis, and regularly use antithrombotic therapy and nonsteroidal anti-inflammatory agents. Additional risk factors include obesity and a history of diabetes mellitus. Most bleeds are self-limited and stop spontaneously with bowel rest. The lifetime risk of rebleeding is 25%.

Initial localization of diverticular bleeding may include colonoscopy, multiplanar computed tomography (CT) angiogram, or nuclear medicine tagged red cell scan. If the patient is stable, ongoing bleeding is best managed by angiography. If mesenteric angiography can localize the bleeding site, the vessel can be occluded successfully with a coil in 80% of cases. The patient can then be followed closely with repetitive colonoscopy, if necessary, looking for evidence of colonic ischemia. However, with highly selective coil embolization, the rate of colonic ischemia is <10%, and the risk of acute rebleeding is <25%. Long-term results (40 months) indicate that >50% of patients with acute diverticular bleeds treated with highly selective angiography have had definitive treatment. Alternatively, colonoscopic ligation with banding or placement of a detachable snare has been shown in a recent meta-analysis to be an effective way to obtain hemostasis if the bleeding site can be localized. Ligation has been shown to prevent rebleeding or the requirement of emergent surgery. In the event that these measures fail to achieve hemostasis, a segmental resection of the colon may be undertaken. This may be advantageous in patients on chronic anticoagulation and immunosuppression as delayed bleeding and perforation have been reported in this subpopulation.

If the patient is unstable or has had a 6-unit bleed within 24 h, current recommendations are that surgery should be performed. If the bleeding has been localized, a segmental resection can be performed. If the site of bleeding has not been definitively identified, a subtotal colectomy may be required. In patients without severe comorbidities, surgical resection can be performed with a primary anastomosis. A higher anastomotic leak rate has been reported in patients who received >10 units of blood.

Presentation, Evaluation, and Staging of Diverticulitis Acute uncomplicated diverticulitis (also known as symptomatic uncomplicated diverticular disease [SUDD]) characteristically presents with fever, anorexia, left lower quadrant abdominal pain, and obstipation (Table 328-1). In <25% of cases, patients may present with generalized peritonitis indicating the presence of a diverticular perforation. If a pericolonic abscess has formed, the patient may have abdominal distention and signs of localized peritonitis. Laboratory investigations will demonstrate a leukocytosis. Rarely, a patient may present with an air-fluid level in the left lower quadrant on plain abdominal film. This is a giant diverticulum of the sigmoid colon and is managed with resection to avoid impending perforation.

The diagnosis of diverticulitis is best made on a contrast-enhanced abdominal and pelvic CT scan demonstrating the following findings: sigmoid diverticula, thickened colonic wall >4 mm, and inflammation within the pericolic fat with or without the collection of contrast

TABLE 328-1 Presentation of Diverticular Disease

Uncomplicated Diverticular Disease—75%

Absdominal pain
Fever
Leukocytosis
Anorexia/obstipation

Complicated Diverticular Disease—25%

Abscess 16%
Perforation 10%
Stricture 5%
Fistula 2%

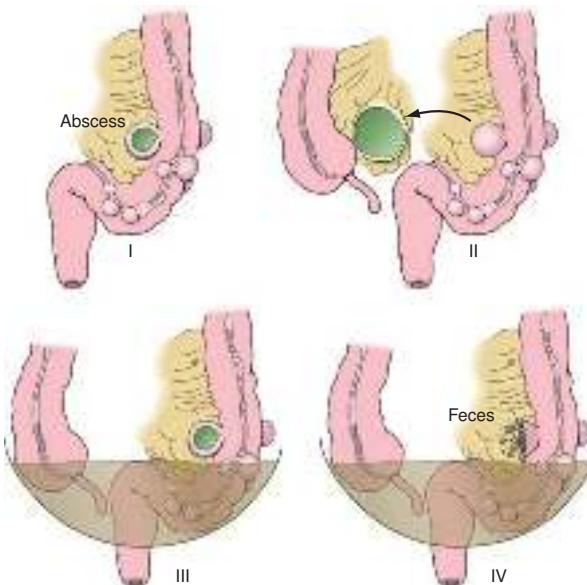


FIGURE 328-2 Hinckley classification of diverticulitis. Stage I: Perforated diverticulitis with a confined paracolic abscess. Stage II: Perforated diverticulitis that has closed spontaneously with distant abscess formation. Stage III: Noncommunicating perforated diverticulitis with fecal peritonitis (the diverticular neck is closed off, and therefore, contrast will not freely expel on radiographic images). Stage IV: Perforation and free communication with the peritoneum, resulting in fecal peritonitis.

material or fluid. In up to 20% of patients, an abdominal abscess may be present. Symptoms of irritable bowel syndrome (Chap. 327) may mimic those of diverticulitis. Therefore, suspected diverticulitis that does not meet CT criteria or is not associated with a leukocytosis or fever is not diverticular disease. Other conditions that can mimic diverticular disease include an ovarian cyst, endometriosis, acute appendicitis, and pelvic inflammatory disease.

Although the benefit of colonoscopy in the evaluation of patients with diverticular disease has been called into question, its use is still considered important in the exclusion of colorectal cancer. The parallel epidemiology of colorectal cancer and diverticular disease provides enough concern for an endoscopic evaluation before operative management. Therefore, a colonoscopy should be performed ~6 weeks after an attack of diverticular disease.

Complicated diverticular disease is defined as diverticular disease associated with an abscess or perforation and less commonly with a fistula (Table 328-1). Perforated diverticular disease is staged using the Hinckley classification system (Fig. 328-2). This staging system was developed to predict outcomes following the surgical management of complicated diverticular disease. In recent years, the Hinckley staging system has been modified to include the development of a phlegmon or early abscess (Hinckley stage Ia). A pericolic abscess is then considered Hinckley stage Ib. In complicated diverticular disease with fistula formation, common locations include cutaneous, vaginal, or vesicle fistulas. These conditions present with either passage of stool through the skin or vagina or the presence of air in the urinary stream (pneumaturia). Colovaginal fistulas are more common in women who have undergone a hysterectomy.

TREATMENT

Diverticular Disease

MEDICAL MANAGEMENT

Asymptomatic diverticular disease discovered on imaging studies or at the time of colonoscopy is best managed by lifestyle changes. Although the data regarding dietary risks and symptomatic diverticular disease are limited (Table 328-2), patients may benefit from

TABLE 328-2 The Use of Fiber in the Management of Diverticular Disease (DD)

JOURNAL, STUDY YEAR	PATIENTS (N)	INTERVENTION	STUDY LENGTH	FINDINGS
Lancet, 1977	18	Wheat or bran crisp bread	3 months	Significant reduction of symptoms score
BMJ, 1981	58	Bran, ispaghula, placebo	16 weeks	No difference
J Gastroenterol, 1977	30	Methylcellulose	3 months	Significant reduction in symptoms
BMJ, 2011	47,033	Vegetarian vs nonvegetarian	11.6 years	Vegetarians had a 31% lower risk of DD
Gastroenterology, 2012	2104	Fiber consumption	12 years	Fiber associated with great risk of DD
JAMA, 2008	47,288	Nut, corn, popcorn consumption	18 years	Higher nut, corn, and popcorn had lower risk of recurrence
Ann R Coll Surg Engl, 1985	56	Fiber consumption	66 months	Higher fiber associated with 19% reduction in symptom recurrence

Source: Modified from A Turis et al: Review article: The pathophysiology and medical management of diverticulosis and diverticular disease of the colon. Aliment Pharmacol Ther 42:664, 2015.

a fiber-enriched diet that includes 30 g of fiber each day. Supplementary fiber products such as Metamucil, Fibercon, or Citrucel are useful. The use of fiber decreases colonic transit time and, therefore, prevents increased intraluminal pressure leading to the development of diverticulosis. The incidence of complicated diverticular disease appears to also be increased in patients who smoke and are obese. Therefore, patients should be encouraged to refrain from smoking and to join a weight loss program. The historical recommendation to avoid eating nuts is based on no more than anecdotal data.

SUDD with confirmation of inflammation and infection within the colon should be treated initially with bowel rest. The routine use of antibiotics in uncomplicated diverticular disease does not reduce time to symptom resolution or risk of complications or recurrence. There is ample evidence establishing the safety of treating SUDD without antibiotics. The AVOD trial randomly assigned 623 inpatients with CT-confirmed uncomplicated left-sided diverticulitis to receive intravenous fluids alone or intravenous fluids and antibiotics and found no differences between the treatment groups in terms of time to recovery, the development of complicated diverticular disease, and recurrence. The DIABOLO trial from the Dutch Diverticular Disease Collaborative Study Group compared the efficacy of treating patients with their first episode of sigmoid diverticulitis with antibiotics or outpatient observation. A total of 528 patients with CT-proven uncomplicated diverticulitis were randomized to either a 10-day course of amoxicillin/clavulanic acid or observation in an outpatient setting. This study demonstrated that antibiotics made no difference in symptom duration or management, and the results favored observation over antibiotic therapy for uncomplicated diverticulitis (SUDD). Currently, the practice guidelines of the American Society of Colon and Rectal Surgery state that "selected patients with uncomplicated diverticulitis can be treated without antibiotics." Hospitalization for acute diverticulitis is recommended if the patient is unable to take oral therapy, is affected by several comorbidities, fails to improve with outpatient therapy, or has complicated diverticulitis. Nearly 75% of patients hospitalized for acute diverticulitis will respond to nonoperative treatment with a suitable antimicrobial regimen. The current recommended antimicrobial coverage is a third-generation cephalosporin or ciprofloxacin and metronidazole targeting aerobic gram-negative rods and anaerobic bacteria. Unfortunately, these agents do not cover enterococci, and the addition of ampicillin to this regimen for nonresponders is recommended. Alternatively, single-agent therapy with a third-generation penicillin such as IV piperacillin or oral penicillin/clavulanic acid may be effective. The usual course of antibiotics is 7–10 days, although this length of time is being investigated. Patients should remain on a limited diet until their pain resolves.

Once the acute attack has resolved, the mainstay medical management of diverticular disease to prevent symptoms has evolved. The risk of recurrent hospitalization following an episode of acute diverticular disease is 11%. Established risk factors for symptomatic recurrence include younger age, the formation of a diverticular abscess, more frequent attacks (>2 per year), multimorbidity, obesity, and smoking. Prevention strategies may include smoking

cessation and weight loss. Diverticular disease is now considered a functional bowel disorder associated with low-grade inflammation. The use of anti-inflammatory medications (mesalazine) in randomized clinical trials has shown them to be beneficial at reducing symptoms and disease recurrence in patients with SUDD. However, when objective signs of inflammation such as C-reactive protein and computerized imaging are taken into consideration, no benefit for the use of mesalazine has been shown.

Treatment strategies targeting dysbiosis in diverticular disease have also been evaluated using polymerase chain reaction (PCR) on stool specimens. Stool samples from consumers of a high-fiber diet have different bacterial content than stool samples from consumers of a low-fiber, high-fat diet. Probiotics are increasingly used by gastroenterologists for multiple bowel disorders and may prevent recurrence of diverticulitis. Specifically, probiotics containing *Lactobacillus acidophilus* and *Bifidobacterium* strains may be beneficial; however, a recent systematic review was unable to show any benefit to the use of probiotics alone. The addition of fiber or mesalazine with probiotics has been shown to maintain remission. Rifaximin (a poorly absorbed broad-spectrum antibiotic), when compared to fiber alone for the treatment of SUDD, is associated with 30% less frequent recurrent symptoms from uncomplicated diverticular disease.

SURGICAL MANAGEMENT

Preoperative risk factors influencing postoperative mortality rates include higher American Society of Anesthesiologists (ASA) physical status class (Table 328-3) and preexisting organ failure. In patients who are low risk (ASA P1 and P2), surgical therapy can be offered to those who do not rapidly improve on medical therapy. For uncomplicated diverticular disease, medical therapy can be continued beyond two attacks without an increased risk of perforation requiring a colostomy. However, patients on immunosuppressive therapy, in chronic renal failure, or with a collagen-vascular disease have a fivefold greater risk of perforation during recurrent attacks. A multicentered randomized clinical trial (DIRECT trial) comparing surgery with conservative management for recurrent SUDD demonstrated that elective surgical resection was associated with an improved quality of life and was more cost-effective at 5 years following resection as compared to conservative management. Surgical therapy is indicated in all low-surgical-risk patients with complicated diverticular disease.

The goals of surgical management of diverticular disease include controlling sepsis, eliminating complications such as fistula or obstruction, removing the diseased colonic segment, and restoring intestinal continuity. These goals must be obtained while minimizing morbidity rate, length of hospitalization, and cost in addition to maximizing survival and quality of life. Table 328-4 lists the operations most commonly indicated based on the Hinchey classification and the predicted postoperative outcomes. The current options for uncomplicated diverticular disease include an open or a laparoscopic resection of the diseased area with reanastomosis to the rectosigmoid. Preservation of portions of the sigmoid colon may lead to early recurrence of the disease. The benefits of laparoscopic

TABLE 328-3 American Society of Anesthesiologists Physical Status Classification System

P1	A normal healthy patient
P2	A patient with mild systemic disease
P3	A patient with severe systemic disease
P4	A patient with severe systemic disease that is a constant threat to life
P5	A moribund patient who is not expected to survive without the operation
P6	A declared brain-dead patient whose organs are being removed for donor purposes

resection over open surgical techniques include early discharge (by at least 1 day), less narcotic use, less postoperative complications, and an earlier return to work.

The options for the surgical management of complicated diverticular disease (Fig. 328-3) include the following open or laparoscopic procedures: (1) proximal diversion of the fecal stream with an ileostomy or colostomy and sutured omental patch with drainage, (2) resection with colostomy and mucous fistula or closure of distal bowel with formation of a Hartmann's pouch (Hartmann's procedure), (3) resection with anastomosis (coloproctostomy), or (4) resection with anastomosis and diversion (coloproctostomy with loop ileostomy or colostomy). (5) Laparoscopic technique of washout and drainage without diversion has been described for Hinchey III patients; however, a threefold increased risk of recurrent peritonitis requiring reoperation with washout alone has been reported.

Patients with Hinchey stage Ia are managed with antibiotic therapy only followed by resection with anastomosis at 6 weeks. Patients with Hinchey stages Ib and II disease are managed with percutaneous drainage followed by resection with anastomosis about 6 weeks later. Current guidelines put forth by the American Society of Colon and Rectal Surgeons suggest, in addition to antibiotic therapy, CT-guided percutaneous drainage of diverticular abscesses that are >3 cm and have a well-defined wall. Abscesses that are <5 cm may resolve with antibiotic therapy alone. Contraindications to percutaneous drainage are no percutaneous access route, pneumoperitoneum, and fecal peritonitis. Drainage of a diverticular abscess is associated with a 20–25% failure rate. Urgent operative intervention is undertaken if percutaneous drainage fails and patients develop generalized peritonitis, and most will need to be managed with a Hartmann's procedure (resection of the sigmoid colon with end colostomy and rectal stump). In selected cases, nonoperative therapy may be considered. In one nonrandomized study, nonoperative management of isolated paracolic abscesses (Hinchey stage I) was associated with only a 20% recurrence rate at 2 years.

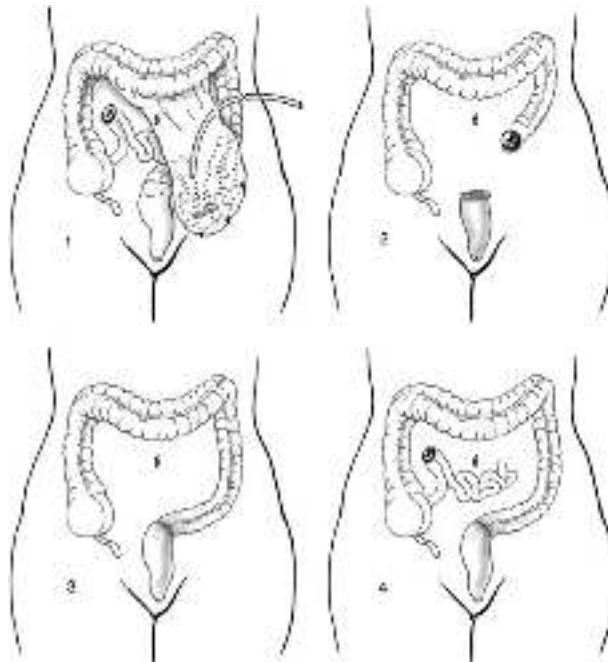


FIGURE 328-3 Methods of surgical management of complicated diverticular disease. 1. Drainage, omental pedicle graft, and proximal diversion. 2. Hartmann's procedure. 3. Sigmoid resection with coloproctostomy. 4. Sigmoid resection with coloproctostomy and proximal diversion.

More than 80% of patients with distant abscesses (Hinchey stage II) required surgical resection for recurrent symptoms.

The management of Hinchey stage III disease is under debate. In this population of patients, no fecal peritonitis is present, and it is presumed that the perforation has sealed. Historically, Hinchey stage III has been managed with a Hartmann's procedure or with primary anastomosis and proximal diversion. Several studies have examined short- and long-term outcomes for laparoscopic peritoneal lavage to remove the peritoneal contamination and place drainage catheters should a communication to the bowel still exist. However, this procedure has been associated with an increased risk of requiring reoperation for ongoing peritonitis. Overall, ostomy rates are lower with the use of laparoscopic peritoneal lavage. No anastomosis of any type should be attempted in Hinchey stage IV disease or in the presence of fecal peritonitis. A limited approach to these patients is associated with a decreased mortality rate.

TABLE 328-4 Outcome Following Surgical Therapy for Complicated Diverticular Disease Based on Modified Hinchey Staging

HINCHY STAGE	OPERATIVE PROCEDURE	ANASTOMOTIC LEAK RATE, %	OVERALL MORBIDITY RATE, %
Ia (pericolic phlegmon)	Laparoscopic or open colon resection	43	15
Ib (pericolic abscess)	Percutaneous drainage followed by laparoscopic or open colon resection	3	15
II	Percutaneous drainage followed by laparoscopic or open colon resection +/- proximal diversion with an ostomy	3	15
III	Laparoscopic washout and drainage or Laparoscopic or open resection with proximal diversion (ostomy) or Hartmann's procedure	3	30% risk of peritonitis requiring reoperation if no resection is performed. Overall morbidity 50% Overall mortality 15%
IV	Hartmann's procedure or Washout with proximal diversion	—	Overall morbidity 50% Overall mortality 15%

Recurrent Symptoms Recurrent abdominal symptoms following surgical resection for diverticular disease occur in 10% of patients. Recurrent diverticular disease develops in patients following inadequate surgical resection. A retained segment of diseased rectosigmoid colon is associated with twice the incidence of recurrence. The presence of irritable bowel syndrome may also cause recurrence of initial symptoms. Patients undergoing surgical resection for presumed diverticulitis and symptoms of chronic abdominal cramping and irregular loose bowel movements consistent with irritable bowel syndrome have poorer functional outcomes.

COMMON DISEASES OF THE ANORECTUM

RECTAL PROLAPSE PROCIDENTIA

Incidence and Epidemiology Rectal prolapse is six times more common in women than in men. The incidence of rectal prolapse peaks in women >60 years. Women with rectal prolapse have a higher incidence of associated pelvic floor disorders including urinary incontinence, rectocele, cystocele, and enterocele. About 20% of children with rectal prolapse will have cystic fibrosis. All children presenting with prolapse should undergo a sweat chloride test. Less common associations include Ehlers-Danlos syndrome, solitary rectal ulcer syndrome, congenital hypothyroidism, Hirschsprung's disease, dementia, cognitively impaired, and schizophrenia.

Anatomy and Pathophysiology Rectal prolapse (procidentia) is a circumferential, full-thickness protrusion of the rectal wall through the anal orifice. It is often associated with a redundant sigmoid colon, pelvic laxity, and a deep rectovaginal septum (pouch of Douglas). Initially, rectal prolapse was felt to be the result of early internal rectal intussusception, which occurs in the upper to mid rectum. This was considered to be the first step in an inevitable progression to full-thickness external prolapse. However, only 1 of 38 patients with internal prolapse followed for >5 years developed full-thickness prolapse. Others have suggested that full-thickness prolapse is the result of damage to the nerve supply to the pelvic floor muscles or pudendal nerves from repeated stretching with straining to defecate. Damage to the pudendal nerves would weaken the pelvic floor muscles, including the external anal sphincter muscles. Bilateral pudendal nerve injury is more significantly associated with prolapse and incontinence than unilateral injury.

Presentation and Evaluation In external prolapse, the majority of patient complaints include anal mass, bleeding per rectum, and poor perianal hygiene. Prolapse of the rectum usually occurs following defecation and will spontaneously reduce or require the patient to manually reduce the prolapse. Constipation occurs in ~30–67% of patients with rectal prolapse. Differing degrees of fecal incontinence occur in 50–70% of patients. Patients with internal rectal prolapse will present with symptoms of both constipation and incontinence. Other associated findings include outlet obstruction (anismus) in 30%, colonic inertia in 10%, and solitary rectal ulcer syndrome in 12%.

Office evaluation is best performed after the patient has been given an enema, which enables the prolapse to protrude. An important distinction should be made between full-thickness rectal prolapse and isolated mucosal prolapse associated with hemorrhoidal disease (Fig. 328-4). Mucosal prolapse is known for radial grooves rather than circumferential folds around the anus and is due to increased laxity of the connective tissue between the submucosa and underlying muscle of the anal canal. The evaluation of prolapse should also include cystoproctography and colonoscopy. These examinations evaluate for associated pelvic floor disorders and rule out a malignancy or a polyp as the lead point for prolapse. If rectal prolapse is associated with chronic constipation, the patient should undergo a defecating proctogram and a sitzmark study. This will evaluate for the presence of anismus or colonic inertia. Anismus is the result of attempting to defecate against a closed pelvic floor and is also known as *nonrelaxing puborectalis*. This can be seen when straightening of the rectum fails to occur on fluoroscopy while the patient is attempting to defecate. In colonic inertia, a sitzmark study will demonstrate retention of >20% of

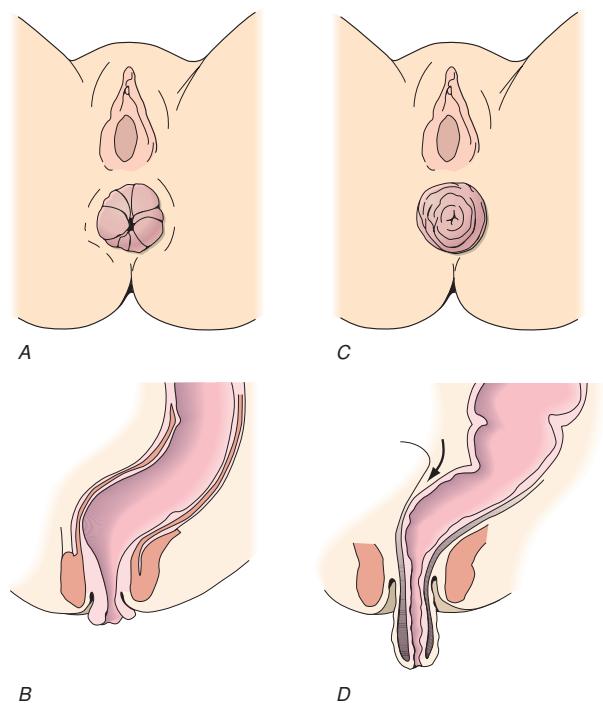


FIGURE 328-4 Degrees of rectal prolapse. Mucosal prolapse only (A, B, sagittal view). Full-thickness prolapse associated with redundant rectosigmoid and deep pouch of Douglas (C, D, sagittal view).

markers on abdominal x-ray 5 days after swallowing. For patients with fecal incontinence, endoanal ultrasound and manometric evaluation, including pudendal nerve testing of their anal sphincter muscles, may be performed before surgery for prolapse (see “Fecal Incontinence,” below).

TREATMENT

Rectal Prolapse

The medical approach to the management of rectal prolapse is limited and includes stool-bulking agents or fiber supplementation to ease the process of evacuation. Surgical correction of rectal prolapse is the mainstay of therapy. Two approaches are commonly considered: transabdominal and transperineal. Transabdominal approaches have been associated with lower recurrence rates, but some patients with significant comorbidities are better served by a transperineal approach.

Common transperineal approaches include a transanal proctectomy (Altmeyer procedure), mucosal proctectomy (Delorme procedure), or placement of a Tirsch wire encircling the anus. The goal of the transperineal approach is to remove the redundant rectosigmoid colon. Common transabdominal approaches include presacral suture or mesh rectopexy (Ripstein) with (Frykman-Goldberg) or without resection of the redundant sigmoid. Colon resection, in general, is reserved for patients with constipation and outlet obstruction. Ventral rectopexy is an effective method of abdominal repair of full-thickness prolapse that does not require sigmoid resection (see description below). This repair may have improved functional results over other abdominal repairs. Transabdominal procedures can be performed effectively with laparoscopic and, more recently, robotic techniques without increased incidence of recurrence. The goal of the transabdominal approach is to restore normal anatomy by removing redundant bowel and reattaching the supportive tissue of the rectum to the presacral fascia. The final alternative is abdominal proctectomy with end-sigmoid colostomy. If total colonic inertia is present, as defined by a history of

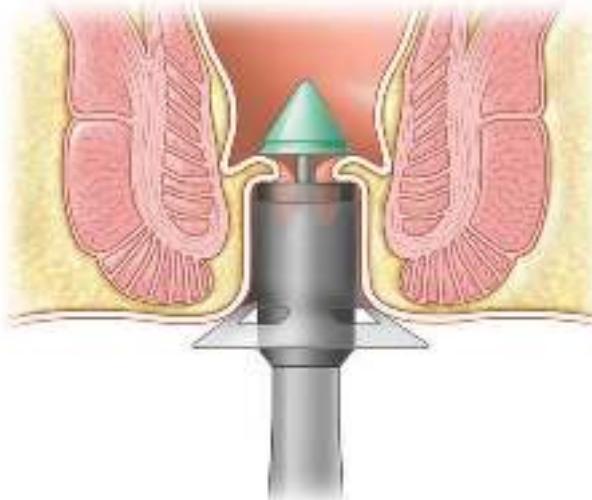


FIGURE 328-5 Stapled transanal rectal resection. Schematic of placement of the circular stapling device.

constipation and a positive sitzmark study, a subtotal colectomy with an ileosigmoid or rectal anastomosis may be required at the time of rectopexy.

Previously, the presence of internal rectal prolapse identified on imaging studies has been considered a nonsurgical disorder, and biofeedback was recommended. However, only one-third of patients will have successful resolution of symptoms from biofeedback. Two surgical procedures more effective than biofeedback are the stapled transanal rectal resection (STARR) and the laparoscopic ventral rectopexy (LVR). The STARR procedure (Fig. 328-5) is performed through the anus in patients with internal prolapse. A circular stapling device is inserted through the anus; the internal prolapse is identified and ligated with the stapling device. LVR (Fig. 328-6) is performed through an abdominal approach. An opening in the peritoneum is created on the left side of the rectosigmoid junction, and this opening continues down anterior on the rectum into the pouch of Douglas. No rectal mobilization is performed, thus avoiding any autonomic nerve injury. Mesh is secured to the anterior and lateral portion of the rectum, the vaginal fornix, and the sacral promontory, allowing for closure of the rectovaginal septum and correction of the internal prolapse. In both procedures, recurrence at 1 year was low (<10%), and symptoms improved in more than three-fourths of patients.

FECAL INCONTINENCE

Incidence and Epidemiology Fecal incontinence is the involuntary passage of fecal material for at least 1 month in an individual with a developmental age of at least 4 years. The prevalence of fecal incontinence in the United States is 0.5–11%. The majority of patients are women and aged >65. A higher incidence of incontinence is seen among parous women. One-half of patients with fecal incontinence also suffer from urinary incontinence. The majority of incontinence is a result of obstetric injury to the pelvic floor, either while carrying a fetus or during the delivery. An anatomic sphincter defect may occur in up to 32% of women following childbirth regardless of visible damage to the perineum. Risk factors at the time of delivery include prolonged labor, the use of forceps, and the need for an episiotomy. Symptoms of incontinence can present two or more decades after obstetric injury. Medical conditions known to contribute to the development of fecal incontinence are listed in Table 328-5.

Anatomy and Pathophysiology The anal sphincter complex is made up of the internal and external anal sphincter. The internal sphincter is smooth muscle and a continuation of the circular fibers of the rectal wall. It is innervated by the intestinal myenteric plexus and

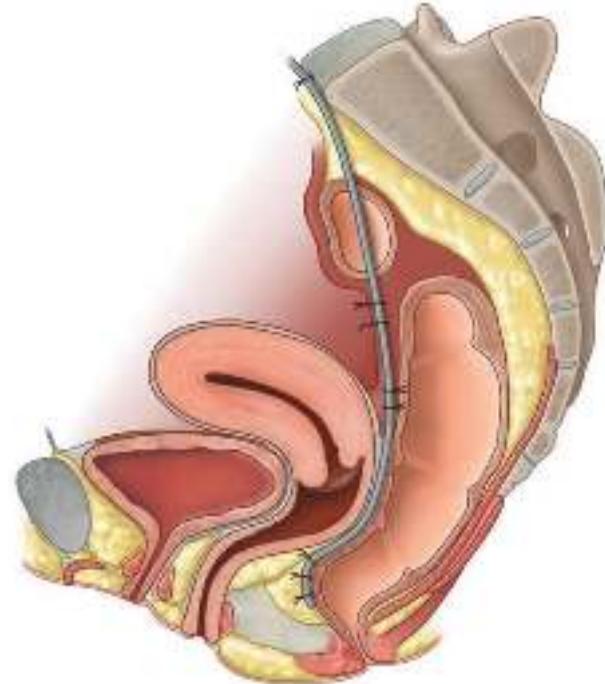


FIGURE 328-6 Laparoscopic ventral rectopexy (LVR). To reduce the internal prolapse and close any rectovaginal septal defect, the pouch of Douglas is opened and mesh is secured to the anterolateral rectum, vaginal fornix, and sacrum. (Reproduced with permission from A D'Hoore et al: Long-term outcome of laparoscopic ventral rectopexy for total rectal prolapse. *BJS* 91:1500, 2004.)

is therefore not under voluntary control. The external anal sphincter is formed in continuation with the levator ani muscles and is under voluntary control. The pudendal nerve supplies motor innervation to the external anal sphincter. Obstetric injury may result in tearing of the muscle fibers anteriorly at the time of the delivery. This results in an obvious anterior defect on endoanal ultrasound. Injury may also be the result of stretching of the pudendal nerves during pregnancy or delivery of the fetus through the birth canal.

Presentation and Evaluation Patients may suffer with varying degrees of fecal incontinence. Minor incontinence includes incontinence to flatus and occasional seepage of liquid stool. Major incontinence is frequent inability to control solid waste. As a result of fecal incontinence, patients suffer from poor perianal hygiene. Beyond the immediate

TABLE 328-5 Medical Conditions That Contribute to Symptoms of Fecal Incontinence

Neurologic Disorders
<ul style="list-style-type: none"> Dementia Brain tumor Stroke Multiple sclerosis Tabes dorsalis Cauda equina lesions
Skeletal Muscle Disorders
<ul style="list-style-type: none"> Myasthenia gravis Myopathies, muscular dystrophy
Miscellaneous
<ul style="list-style-type: none"> Hypothyroidism Irritable bowel syndrome Diabetes Severe diarrhea Scleroderma

problems associated with fecal incontinence, these patients are often withdrawn and suffer from depression. For this reason, quality-of-life measures are an important component in the evaluation of patients with fecal incontinence.

The evaluation of fecal incontinence should include a thorough history and physical examination including digital rectal examination (DRE). Weak sphincter tone on DRE and loss of the “anal wink” reflex (S1-level control) may indicate a neurogenic dysfunction. Perianal scars may represent surgical injury. Other studies helpful in the diagnosis of fecal incontinence include anal manometry, pudendal nerve terminal motor latency (PNTML), and endoanal ultrasound. Centers that care for patients with fecal incontinence will have an anorectal physiology laboratory that uses standardized methods of evaluating anorectal physiology. Anorectal manometry (ARM) measures resting and squeeze pressures within the anal canal using an intraluminal water-perfused catheter. Current methods of ARM include use of a three-dimensional, high-resolution system with a 12-catheter perfusion system, which allows physiologic delineation of anatomic abnormalities. Pudendal nerve studies evaluate the function of the nerves innervating the anal canal using a finger electrode placed in the anal canal. Stretch injuries to these nerves will result in a delayed response of the sphincter muscle to a stimulus, indicating a prolonged latency. Finally, endoanal ultrasound will evaluate the extent of the injury to the sphincter muscles before surgical repair. Unfortunately, all of these investigations are user-dependent, and very few studies demonstrate that these studies predict outcome following an intervention. Magnetic resonance imaging (MRI) has been used, but its routine use for imaging in fecal incontinence is not well established.

Rarely does a pelvic floor disorder exist alone. The majority of patients with fecal incontinence will have some degree of urinary incontinence. Similarly, fecal incontinence is a part of the spectrum of pelvic organ prolapse. For this reason, patients may present with symptoms of obstructed defecation as well as fecal incontinence. Careful evaluation including dynamic MRI or cinedefecography should be performed to search for other associated defects. Surgical repair of incontinence without attention to other associated defects may decrease the success of the repair.

TREATMENT

Fecal Incontinence

Medical management of fecal incontinence includes strategies to bulk up the stool, which help in increasing fecal sensation. These include fiber supplementation, loperamide, diphenoxylate, and bile acid binders. These agents harden the stool and delay frequency of bowel movements and are helpful in patients with minimal to mild symptoms. Furthermore, patients can be offered a form of physical therapy called biofeedback. This therapy helps strengthen the external sphincter muscle while training the patient to relax with defecation to avoid unnecessary straining and further injury to the sphincter muscles. Biofeedback has had variable success and is dependent on the motivation of the patient. At a minimum, biofeedback is risk-free. Most patients will have some improvement. For this reason, it should be incorporated into the initial recommendation to all patients with fecal incontinence.

Historically, the “gold standard” for the treatment of fecal incontinence with an isolated sphincter defect has been the overlapping sphincteroplasty. The external anal sphincter muscle and scar tissue, as well as any identifiable internal sphincter muscle, are dissected free from the surrounding adipose and connective tissue, and then an overlapping repair is performed in an attempt to rebuild the muscular ring and restore its function. However, long-term results following overlapping sphincteroplasty have been poor, with a 50% failure rate over 5 years.

Alternative therapies such as sacral nerve stimulation (SNS), collagen-enhancing injectables, and magnetic “Fenix” ring are other options. SNS is an adaptation of a procedure developed for the management of urinary incontinence. SNS is ideally suited for

patients with intact but weak anal sphincters. A temporary nerve stimulator is placed on the third sacral nerve. If there is at least a 50% improvement in symptoms, a permanent nerve stimulator is placed under the skin. Long-term results for sacral stimulation have been promising, with nearly 80% of patients having a reduction in incontinence episodes by at least 50%. This reduction has been sustainable in studies out to 5 years. Collagen-enhancing injectables have been around for several years. More than 50% of incontinent patients treated with nonanimal stabilized hyaluronic acid (NASHA/DX) achieved a 50% reduction in incontinence episodes, and these results were sustainable up to 2 years. Currently, this injectable is not universally available. The Fenix is a magnetic ring that is implanted around the anal sphincter muscles. Its long-term outcomes are still being studied, and it is currently only available for compassionate use.

Finally, the use of stem cells to increase the bulk of the sphincter muscles is currently being tested. Stem cells can be harvested from the patient’s own muscle, grown, and then implanted into their sphincter complex. Concern for cost and the need for an additional procedure dampen enthusiasm. Trial results are awaited.

■ HEMORRHOIDAL DISEASE

Incidence and Epidemiology Symptomatic hemorrhoids affect >1 million individuals in the Western world per year. The prevalence of hemorrhoidal disease is not selective for age or sex. However, age is known to be a risk factor. The prevalence of hemorrhoidal disease is less in underdeveloped countries. The typical low-fiber, high-fat Western diet is associated with constipation and straining and the development of symptomatic hemorrhoids.

Anatomy and Pathophysiology Hemorrhoidal cushions are a normal part of the anal canal. The vascular structures contained within this tissue aid in continence by preventing damage to the sphincter muscle. Three main hemorrhoidal complexes traverse the anal canal—the left lateral, the right anterior, and the right posterior. Engorgement and straining lead to prolapse of this tissue into the anal canal. Over time, the anatomic support system of the hemorrhoidal complex weakens, exposing this tissue to the outside of the anal canal where it is susceptible to injury. Hemorrhoids are commonly classified as external or internal. External hemorrhoids originate below the dentate line and are covered with squamous epithelium and are associated with an internal component. External hemorrhoids are painful when thrombosed. Internal hemorrhoids originate above the dentate line and are covered with mucosa and transitional zone epithelium and represent the majority of hemorrhoids. The standard classification of hemorrhoidal disease is based on the progression of the disease from their normal internal location to the prolapsing external position (Table 328-6).

TABLE 328-6 The Staging and Treatment of Hemorrhoids

STAGE	DESCRIPTION OF CLASSIFICATION	TREATMENT
I	Enlargement with bleeding	Fiber supplementation Short course of cortisone suppository Sclerotherapy Infrared coagulation
II	Protrusion with spontaneous reduction	Fiber supplementation Short course of cortisone suppository Sclerotherapy Infrared coagulation
III	Protrusion requiring manual reduction	Fiber supplementation Short course of cortisone suppository Rubber band ligation Operative hemorrhoidectomy
IV	Irreducible protrusion	Fiber supplementation Cortisone suppository Operative hemorrhoidectomy

2504 Presentation and Evaluation Patients commonly present to a physician for two reasons: bleeding and protrusion. Pain is less common than with fissures and, if present, is described as a dull ache from engorgement of the hemorrhoidal tissue. Severe pain may indicate a thrombosed hemorrhoid. Hemorrhoidal bleeding is described as painless bright red blood seen either in the toilet or upon wiping. Occasional patients can present with significant bleeding, which may be a cause of anemia; however, the presence of a colonic neoplasm must be ruled out in anemic patients. Patients who present with a protruding mass complain about inability to maintain perianal hygiene and are often concerned about the presence of a malignancy.

The diagnosis of hemorrhoidal disease is made on physical examination. Inspection of the perianal region for evidence of thrombosis or excoriation is performed, followed by a careful digital examination. Anoscopy is performed paying particular attention to the known position of hemorrhoidal disease. The patient is asked to strain. If this is difficult for the patient, the maneuver can be performed while sitting on a toilet. The physician is notified when the tissue prolapses. It is important to differentiate the circumferential appearance of a full-thickness rectal prolapse from the radial nature of prolapsing hemorrhoids (see “Rectal Prolapse,” above). The stage and location of the hemorrhoidal complexes are defined.

TREATMENT

Hemorrhoidal Disease

The treatment for bleeding hemorrhoids is based on the stage of the disease (Table 328-6). In all patients with bleeding, the possibility of other causes must be considered. In young patients without a family history of colorectal cancer, the hemorrhoidal disease may be treated first and a colonoscopic examination performed if the bleeding continues. Older patients who have not had colorectal cancer screening should undergo colonoscopy or flexible sigmoidoscopy.

With rare exceptions, the acutely thrombosed hemorrhoid can be excised within the first 72 h by performing an elliptical excision. Sitz baths, fiber, and stool softeners are prescribed. Additional therapy for bleeding hemorrhoids includes the office procedures of rubber band ligation, infrared coagulation, and sclerotherapy. Sensation begins at the dentate line; therefore, all procedures can be performed without discomfort either endoscopically or in the office. Bands are placed around the engorged tissue, causing ischemia and fibrosis. This aids in fixing the tissue proximally in the anal canal. Patients may complain of a dull ache for 24 h following band application. During sclerotherapy, 1–2 mL of a sclerosant (usually sodium tetradecyl sulfate) is injected using a 25-gauge needle into the submucosa of the hemorrhoidal complex. Care must be taken not to inject the anal canal circumferentially, or stenosis may occur.

For surgical management of hemorrhoidal disease, excisional hemorrhoidectomy with sharp dissection or using a ligator, transhemorrhoidal dearterialization (THD), or stapled hemorrhoidectomy (“the procedure for prolapse or hemorrhoids” [PPH]) is the procedure of choice. All surgical methods of management are equally effective in the treatment of symptomatic third- and fourth-degree hemorrhoids. However, because the sutured hemorrhoidectomy involves the removal of redundant tissue down to the anal verge, unpleasant anal skin tags are removed as well. The stapled hemorrhoidectomy is associated with less discomfort; however, this procedure does not remove anal skin tags, and an increased number of complications are associated with use of the stapling device. THD uses ultrasound guidance to ligate the blood supply to the anal tissue, hence reducing hemorrhoidal engorgement. No procedures on hemorrhoids should be done in patients who are immunocompromised or who have active proctitis. Furthermore, emergent hemorrhoidectomy for bleeding hemorrhoids is associated with a higher complication rate.

Acute complications associated with the treatment of hemorrhoids include pain, infection, recurrent bleeding, and urinary retention. Care should be taken to place bands properly and to avoid overhydration in patients undergoing operative hemorrhoidectomy. Late complications include fecal incontinence as a result of injury to the sphincter during the dissection. Anal stenosis may develop from overzealous excision, with loss of mucosal skin bridges for reepithelialization. Finally, an *ectropion* (prolapse of rectal mucosa from the anal canal) may develop. Patients with an ectropion complain of a “wet” anus as a result of inability to prevent soiling once the rectal mucosa is exposed below the dentate line.

ANORECTAL ABSCESS

Incidence and Epidemiology The development of a perianal abscess is more common in men than women by a ratio of 3:1. The peak incidence is in the third to fifth decade of life. Perianal pain associated with the presence of an abscess accounts for 15% of office visits to a colorectal surgeon. The disease is more prevalent in immunocompromised patients such as those with diabetes, hematologic disorders, or inflammatory bowel disease (IBD) and persons who are HIV positive. These disorders should be considered in patients with recurrent perianal infections.

Anatomy and Pathophysiology An anorectal abscess is an abnormal fluid-containing cavity in the anorectal region. Anorectal abscess results from an infection involving the glands surrounding the anal canal. Normally, these glands release mucus into the anal canal, which aids in defecation. When stool accidentally enters the anal glands, the glands become infected, and an abscess develops. Anorectal abscesses are perianal in 40–50% of patients, ischiorectal in 20–25%, intersphincteric in 2–5%, and suprarelevator in 2.5% (Fig. 328-7).

Presentation and Evaluation Perianal pain and fever are the hallmarks of an abscess. Patients may have difficulty voiding and have blood in the stool. A prostatic abscess may present with similar complaints, including dysuria. Patients with a prostatic abscess will often have a history of recurrent sexually transmitted diseases. On physical examination, a large fluctuant area is usually readily visible. Routine laboratory evaluation shows an elevated white blood cell count. Diagnostic procedures are rarely necessary unless evaluating a recurrent abscess. A CT scan or MRI has an accuracy of 80% in determining incomplete drainage. If there is a concern about the presence of IBD, a rigid or flexible sigmoidoscopic examination may be done at the time of drainage to evaluate for inflammation within the rectosigmoid region. A more complete evaluation for Crohn’s disease would include a full colonoscopy and small-bowel series.

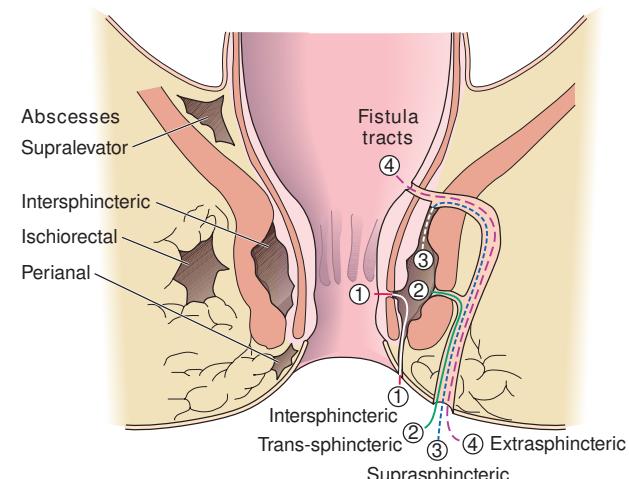


FIGURE 328-7 Common locations of anorectal abscess (left) and fistula in ano (right).

TREATMENT

Anorectal Abscess

As with all abscesses, the “gold standard” is drainage. Office drainage of an uncomplicated anorectal abscess may suffice. A small incision close to the anal verge is made, and a Mallenot drain is advanced into the abscess cavity. For patients who have a complicated abscess or who are diabetic or immunocompromised, drainage should be performed in an operating room under anesthesia. These patients are at greater risk for developing necrotizing fasciitis. The role of antibiotics in the management of anorectal abscesses is limited. Antibiotics are only warranted in patients who are immunocompromised or have prosthetic heart valves, artificial joints, diabetes, or IBD.

■ FISTULA IN ANO

Incidence and Epidemiology The incidence and prevalence of fistulating perianal disease parallel the incidence of anorectal abscess, estimated to be 1 in 10,000 individuals. Some 30–40% of abscesses will give rise to fistula in ano. Although the majority of the fistulas are cryptoglandular in origin, 10% are associated with IBD, tuberculosis, malignancy, and radiation.

Anatomy and Pathophysiology A fistula in ano is defined as a communication of an abscess cavity with an identifiable internal opening within the anal canal. This identifiable opening is most commonly located at the dentate line where the anal glands enter the anal canal. Patients experiencing continuous drainage following the treatment of a perianal abscess likely have a fistula in ano. These fistulas are classified by their relationship to the anal sphincter muscles, with 70% being intersphincteric, 23% transsphincteric, 5% suprasphincteric, and 2% extraspincteric (Fig. 328-7).

Presentation and Evaluation A patient with a fistula in ano will complain of constant drainage from the perianal region associated with a firm mass. The drainage may increase with defecation. Perianal hygiene is difficult to maintain. Examination under anesthesia is the best way to evaluate a fistula. At the time of the examination, anoscopy is performed to look for an internal opening. Diluted hydrogen peroxide will aid in identifying such an opening. In lieu of anesthesia, MRI with an endoanal coil will also identify tracts in 80% of the cases. After drainage of an abscess with insertion of a Mallenot catheter, a fistulogram through the catheter can be obtained in search of an occult fistula tract. Goodsall's rule states that a posterior external fistula will enter the anal canal in the posterior midline, whereas an anterior fistula will enter at the nearest crypt. A fistula exiting >3 cm from the anal verge may have a complicated upward extension and may not obey Goodsall's rule.

TREATMENT

Fistula in Ano

A newly diagnosed draining fistula is best managed with placement of a seton, a vessel loop or silk tie placed through the fistula tract, which maintains the tract open and quiets down the surrounding inflammation that occurs from repeated blockage of the tract. Once the inflammation is less, the exact relationship of the fistula tract to the anal sphincters can be ascertained. A simple fistulotomy can be performed for intersphincteric and low (less than one-third of the muscle) transsphincteric fistulas without compromising continence. For a higher transsphincteric fistula, an anorectal advancement flap in combination with a drainage catheter or fibrin glue may be used. Very long (>2 cm) and narrow tracts respond better to fibrin glue than shorter tracts. Simple ligation of the internal fistula tract (LIFT procedure) has also been used in the management of simple fistula with good success.

Patients should be maintained on stool-bulking agents, nonnarcotic pain medication, and sitz baths following surgery for a fistula.

Early complications from these procedures include urinary retention and bleeding. Later complications are rare (<10%) and include temporary and permanent incontinence. Recurrence is 0–18% following fistulotomy and 20–30% following anorectal advancement flap and the LIFT procedure.

Fistulizing disease of the anus is common in Crohn's disease, and recent evidence has suggested that the use of mesenchymal stem cell therapy may improve healing rates of fistula associated with Crohn's disease. The ADMIRE study examined the use of allogeneic expanded adipose-derived mesenchymal stem cells in the treatment of complex fistula in ano in Crohn's disease. The study included 212 patients randomized to stem cell therapy or placebo. Fistula remission rates at 52 weeks were significantly higher with the use of stem cell therapy over placebo (59 vs 42%, respectively). Currently, there is an international multicenter trial underway.

■ ANAL FISSURE

Incidence and Epidemiology Anal fissures occur at all ages but are more common in the third through the fifth decades. A fissure is the most common cause of rectal bleeding in infancy. The prevalence is equal in males and females. It is associated with constipation, diarrhea, infectious etiologies, perianal trauma, and Crohn's disease.

Anatomy and Pathophysiology Trauma to the anal canal occurs following defecation. This injury occurs in the anterior or, more commonly, posterior anal canal. Irritation caused by the trauma to the anal canal results in an increased resting pressure of the internal sphincter. The blood supply to the sphincter and anal mucosa enters laterally. Therefore, increased anal sphincter tone results in a relative ischemia in the region of the fissure and leads to poor healing of the anal injury. A fissure that is not in the posterior or anterior position should raise suspicion for other causes, including tuberculosis, syphilis, Crohn's disease, and malignancy.

Presentation and Evaluation A fissure can be easily diagnosed on history alone. The classic complaint is pain, which is strongly associated with defecation and is relentless. The bright red bleeding that can be associated with a fissure is less extensive than that associated with hemorrhoids. On examination, most fissures are located in either the posterior or anterior position. A lateral fissure is worrisome because it may have a less benign nature, and systemic disorders should be ruled out. A chronic fissure is indicated by the presence of a hypertrophied anal papilla at the proximal end of the fissure and a sentinel pile or skin tag at the distal end. Often the circular fibers of the hypertrophied internal sphincter are visible within the base of the fissure. If anal manometry is performed, elevation in anal resting pressure and a sawtooth deformity with paradoxical contractions of the sphincter muscles are pathognomonic.

TREATMENT

Anal Fissure

The management of the acute fissure is conservative. Stool softeners for those with constipation, increased dietary fiber, topical anesthetics, glucocorticoids, and sitz baths are prescribed and will heal 60–90% of fissures. Chronic fissures are those present for >6 weeks. These can be treated with modalities aimed at decreasing the anal canal resting pressure including nifedipine ointment applied three times a day and botulinum toxin type A, up to 20 units, injected into the internal sphincter on each side of the fissure. Both treatments are associated with a fissure healing rate of >80%. Surgical management includes anal dilatation and lateral internal sphincterotomy. Usually, one-third of the internal sphincter muscle is divided; it is easily identified because it is hypertrophied. Recurrence rates from medical therapy are higher, but this is offset by a risk of incontinence following sphincterotomy. Lateral internal sphincterotomy may lead to incontinence more commonly in women.

The author thanks Cory Sandore for providing some illustrations for this chapter. Gregory Bulkley, MD, contributed to this chapter in an earlier edition, and some of that material has been retained here.

FURTHER READING

- B AE et al: Surgical interventions and the use of device-aided therapy for the treatment of fecal incontinence and defecatory disorders. *Clin Gastroenterol Hepatol* 15:1844, 2017.
- D Let al: Randomized clinical trial of observation versus antibiotic treatment for a first episode of CT-proven uncomplicated acute diverticulitis (DIABOLO trial). *BJS* 104:52, 2017.
- G M: The evaluation and office management of hemorrhoids for the gastroenterologist. *Curr Gastroenterol Rep* 19:30, 2017.
- P J et al: Long-term efficacy and safety of stem cell therapy (Cx601) for complex perianal fistulas in patients with Crohn's disease. *Gastroenterology* 154:1334, 2018.
- P D, B AE: Management of pelvic floor disorders: Biofeedback and more. *Curr Treat Options Gastroenterol* 12:456, 2014.
- S HV et al: Minimally invasive incision and drainage technique in the treatment of simple subcutaneous abscess in adults. *Am Surg* 83:699, 2017.
- S J et al: Sphincter-sparing anal fistula repair: Are we getting better? *Dis Colon Rectum* 60:1071, 2017.
- T A: Dietary pattern and colonic diverticulosis. *Curr Opin Clin Nutr Metab Care* 20:409, 2017.

most commonly occurs in areas of severe atherosclerotic narrowing at the SMA and celiac artery.

Nonocclusive mesenteric ischemia represents 20% of the cases and is secondary to intestinal ischemia when subjected to acute hemodynamic instability. Hypovolemia, shock, and the use of vasoconstrictive agents (digoxin, α -adrenergic agonists, cocaine) can precipitate ischemia in these patients. It is the most prevalent gastrointestinal disease complicating cardiovascular surgery. The incidence of ischemic colitis following elective aortic repair is 5–9%, and the incidence triples in patients following emergent repair.

Mesenteric venous thrombosis accounts for <10% of cases and is generally precipitated by a hypercoagulable state due to an underlying inherited disorder such as factor V Leiden, prothrombin mutation, protein S deficiency, protein C deficiency, antithrombin deficiency, and antiphospholipid syndrome. It may also occur as a result of acquired thrombophilia in malignancies, hematologic disorders, or use of oral contraceptives.

ANATOMY AND PATHOPHYSIOLOGY

The blood supply to the intestines is supplied by the celiac artery, SMA, and inferior mesenteric artery (IMA) (Fig. 329-1). Extensive collateralization occurs between major mesenteric trunks and branches of the mesenteric arcades. Collateral vessels within the small bowel are numerous and meet within the duodenum and the bed of the pancreas. Collateral vessels within the colon meet at the splenic flexure and descending/sigmoid colon. These areas, which are inherently at risk for decreased blood flow, are known as *Griffiths' point* and *Sudeck's point*, respectively, and are the most common locations for colonic ischemia (Fig. 329-1, shaded areas). The splanchnic circulation can receive up to 30% of the cardiac output. Protective responses to prevent intestinal ischemia include abundant collateralization, autoregulation of blood flow, and the ability to increase oxygen extraction from the blood.

329

Mesenteric Vascular Insufficiency

Maryam Ali Khan, Jaideep Das Gupta,
Mahmoud Malas

INTESTINAL ISCHEMIA

INCIDENCE AND EPIDEMIOLOGY

Intestinal ischemia occurs when splanchnic perfusion fails to meet the metabolic demands of the intestines, resulting in ischemic tissue injury. Mesenteric ischemia affects 2–3 people per 100,000, with an increasing incidence in the aging population. Mortality with acute presentation remains high, between 50 and 80%, and early diagnosis with prompt intervention is crucial in improving clinical outcomes. Intestinal ischemia is further classified as chronic mesenteric ischemia (CMI) or acute mesenteric ischemic (AMI). CMI is secondary to multiple major visceral arterio-occlusive disease, with involvement of the superior mesenteric artery (SMA) most worrisome. AMI is most commonly associated with (1) arterio-occlusive mesenteric ischemia, (2) nonocclusive mesenteric ischemia, and (3) mesenteric venous thrombosis.

CMI is the failure to achieve normal postprandial hyperemic intestinal blood flow. This occurs due to an imbalance between the supply and demand of oxygen metabolites to the intestinal tract similar to cardiac angina. CMI occurs due to significant atherosclerotic disease leading to the narrowing of the SMA and/or celiac artery origins.

AMI is the occurrence of an abrupt cessation of mesenteric blood flow, usually embolic or thrombotic in nature. Approximately 50% of AMI is due to embolus to the mid to distal SMA. Embolus etiology includes atrial fibrillation, recent myocardial infarction, soft atherosclerotic plaque, infective endocarditis, valvular heart disease, and recent cardiac or vascular catheterization. Approximately 25–30% of cases are characterized by an acute-on-chronic thrombosis in patients with preexisting mesenteric atherosclerosis. Thrombotic occlusion

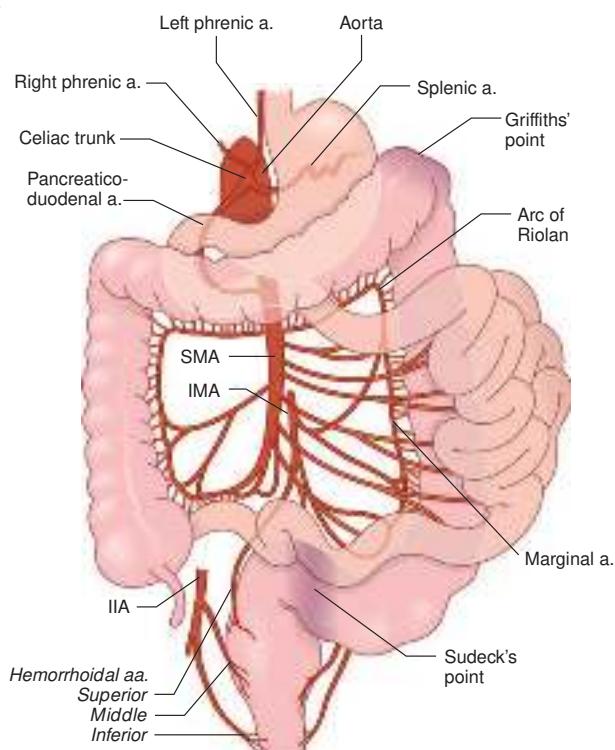


FIGURE 329-1 Blood supply to the intestines includes the celiac artery, superior mesenteric artery (SMA), inferior mesenteric artery (IMA), and branches of the internal iliac artery (IIA). Griffiths' and Sudeck's points, indicated by shaded areas, are watershed areas within the colonic blood supply and common locations for ischemia.

Occlusive ischemia is a result of disruption of blood flow by an embolus or progressive thrombosis in a major artery supplying the intestine. In >75% of cases, emboli originate from the heart and preferentially lodge in the SMA just distal to the origin of the middle colic artery. Progressive thrombosis of typically two of the major vessels supplying the intestine is required for the development of chronic intestinal angina. The involvement of the SMA is most worrisome. Nonocclusive ischemia is disproportionate mesenteric vasoconstriction (arteriolar vasospasm) in response to severe physiologic stress such as shock. If left untreated, early mucosal stress ulceration will progress to full-thickness injury.

PRESENTATION, EVALUATION, AND MANAGEMENT

Patients with CMI typically present with insidious onset of symptoms and classically with recurrent episodes of acute, dull, crampy, post-prandial epigastric pain, which has also been referred to as “intestinal angina.” Weight loss and chronic diarrhea may also be noted. Duration of symptoms is typically 6–12 months. Physical examination often reveals a malnourished patient with other manifestations of atherosclerosis.

Duplex ultrasound has gained popularity as a screening tool for the evaluation of the mesenteric vessels due to high sensitivity and specificity. Mesenteric duplex scan demonstrating a high peak velocity of flow in the SMA is associated with an ~80% positive predictive value of mesenteric ischemia. More significantly, a negative duplex scan virtually precludes the diagnosis of mesenteric ischemia. It is important to perform the test while the patient is fasting because the presence of increased bowel gas prevents adequate visualization of flow disturbances within the vessels or the lack of a vasodilation response to feeding during the test.

The management of CMI includes medical management of the atherosclerotic disease by exercise, cessation of smoking, and antiplatelet and lipid-lowering medications. A full cardiac and vascular evaluation should be performed before intervention on CMI. Before intervention, a CT angiogram is recommended to assess the degree of atherosclerotic disease of the aortic and visceral vessels.

Treatment involves either endovascular or open surgical revascularization and should be individualized based on the patient’s comorbidities and anatomy. Endovascular revascularization involves targeted vessel treatment with visceral stents, with the SMA anatomy being the

key determinant. Open revascularization involves antegrade bypass from the supraceliac aortic or retrograde bypass, typically the common iliac arteries, with a synthetic graft to the targeted vessels, usually the SMA and/or celiac artery. In patients requiring revascularization, the endovascular approach is recommended as first-line therapy. It is especially favorable for short segment stenosis with minimal to moderate calcification or thrombus. Angioplasty with endovascular stenting in the treatment of CMI is associated with an 80% long-term success rate. Open revascularization should be considered in patients with lesions not amenable to endovascular treatment such as severe calcification, longer lesions, small vessel diameter, or failed endovascular interventions.

Acute intestinal ischemia remains one of the most challenging diagnoses. The mortality rate of AMI is >50%. The most significant indicator of survival is the timeliness of diagnosis and treatment. An overview of diagnosis and management of each form of intestinal ischemia is given in Table 329-1.

AMI resulting from an arterial embolus or thrombosis presentation is nonspecific and requires a high index of suspicion for diagnosis. Severe, acute, unremitting abdominal pain strikingly out of proportion to the physical findings is the most common complaint (95%). This may be associated with nausea (44%), vomiting (35%), diarrhea (35%), and blood per rectum (16%). Later findings will demonstrate peritonitis and cardiovascular collapse. Specific clinical features can help differentiate the underlying etiology, whether embolic or thrombotic. Patients with embolic ischemia are typically older adults with an underlying condition that predisposes to embolism such as atrial fibrillation, prior embolic event, or recent infective endocarditis. Thrombotic ischemia typically presents as an acute occlusion in patients with the underlying atherosclerotic disease who may have been previously diagnosed with CMI.

AMI is a surgical emergency, and emergent admission to a monitored bed or intensive care unit is recommended for resuscitation with fluids and administration of broad-spectrum antibiotics in addition to further evaluation. If the diagnosis of intestinal ischemia is being considered, consultation with a surgical service is necessary. Often the decision to operate is made on a high index of suspicion from the history and physical exam despite normal laboratory findings. In patients with suspected AMI, CT angiography with a 1-mm or thinner cut should be used to detect mesenteric arterial occlusive disease most likely due to embolic or thrombotic etiology and is the gold standard.

TABLE 329-1 Overview of the Management of Acute Intestinal Ischemia

CONDITION	KEY TO EARLY DIAGNOSIS	TREATMENT OF UNDERLYING CAUSE	TREATMENT OF SPECIFIC LESION	TREATMENT OF SYSTEMIC CONSEQUENCE
Arterio-occlusive mesenteric ischemia 1. Arterial embolus	Computed tomography (CT) angiography Early laparotomy	Anticoagulation Cardioversion Thrombectomy Broad-spectrum antibiotics	Laparotomy Embolectomy Assess viability and resect nonviable bowel	Anticoagulation Resuscitation Broad-spectrum antibiotics Emergent surgical intervention Assessment of bowel
			Endovascular approach: thrombolysis, angioplasty, and stenting Embolectomy/thrombectomy or vascular bypass Assess viability and resect nonviable bowel	Anticoagulation Resuscitation Broad-spectrum antibiotics Emergent surgical intervention Assessment of bowel
Mesenteric venous thrombosis Venous thrombosis	CT with venous phase	Anticoagulation Resuscitation	Anticoagulation Hypercoagulable workup	Anticoagulation Resuscitation Broad-spectrum antibiotics Support cardiac output Avoid vasoconstrictors
Nonocclusive mesenteric ischemia	Vasospasm: CT Hypoperfusion: CT	Resuscitation Support cardiac output Avoid vasoconstrictors Broad-spectrum antibiotics	Vasospasm: intraarterial vasodilators Hypoperfusion: assess viability and resect dead bowel	Resuscitation Broad-spectrum antibiotics Support cardiac output Avoid vasoconstrictors

Source: Modified from GB Bulkley, in JL Cameron (ed): *Current Surgical Therapy*, 2nd ed. Toronto, BC Decker, 1986.

2508 Additional diagnostic modalities that can be useful in diagnosis but that should not delay surgical therapy include an electrocardiogram (ECG), echocardiogram, and abdominal radiographs. Patients with AMI should be given a heparin bolus and started on a therapeutic heparin drip. Correction of electrolyte abnormalities and empiric broad-spectrum antibiotic therapy should also be initiated immediately.

If the CTA verifies the acute embolic occlusion of SMA, surgical exploration should not be delayed. The goal of operative exploration is to resect the compromised bowel and restore blood supply. The entire length of the small and large bowel beginning at the ligament of Treitz should be evaluated. The SMA artery should be localized, typically at the mesocolon of the transverse colon. A transverse arteriotomy of the SMA should be made with the removal of the embolus with a Fogarty catheter passed in a retrograde and antegrade manner to restore blood flow. In the case of SMA occlusion where the embolus usually lies just proximal to the origin of the middle colic artery, the proximal jejunum is often spared, while the remainder of the small bowel up to the transverse colon may become ischemic. Nonviable bowel should be resected. Questionable bowel should undergo a second-look laparotomy in a 24- to 48-h period. After revascularization, peristalsis and return of pink color of the bowel wall should be observed. Palpation of major arterial mesenteric vessels can be performed, as well as applying a Doppler flowmeter to the antimesenteric border of the bowel wall, but neither is a definitive indicator of viability.

In the assessment of acute-on-chronic mesenteric ischemia, typically involvement of the orifice of the SMA is seen. Therefore, the entire small bowel is compromised. Revascularization using an endovascular, open, and/or hybrid approach should be individualized based on the patient's critical status, comorbidities, and anatomy. Endovascular stenting, suction thrombectomy, and/or thrombolysis catheter should be considered for intervention. The bowel should be evaluated for viability, typically via an exploratory laparotomy.

Noocclusive or vasospastic mesenteric ischemia presents with generalized abdominal pain, anorexia, bloody stools, and abdominal distention. Often these patients are obtunded, and physical findings may not assist in the diagnosis or may be obscured by the underlying etiology. The presence of leukocytosis, metabolic acidosis, and/or lactic acidosis is useful in support of the diagnosis of advanced intestinal ischemia; however, these markers may not be indicative of either reversible ischemia or frank necrosis.

Emergent admission to a monitored bed or intensive care unit is recommended for resuscitation, broad-spectrum antibiotics, and further evaluation. Anticoagulation is not recommended as the goal of resuscitation is to maintain hemodynamics. For select patients, intramesenteric infusion of vasodilators such as papaverine, prostaglandins, and nitroglycerin for reversal of mesenteric ischemia can be used, but resuscitation and the treatment of the underlying pathology should be the priority.

If ischemic colitis is a concern, colonoscopy should be considered to assess the integrity of the colon mucosa. Ischemia of the colonic mucosa is graded as *mild* with minimal mucosal erythema or as *moderate* with pale mucosal ulcerations and evidence of extension to the muscular layer of the bowel wall. *Severe* ischemic colitis presents with severe ulcerations resulting in black or green discoloration of the mucosa, consistent with full-thickness bowel-wall necrosis. Laparoscopy can also be employed for assessment. Ischemic colitis is optimally treated with resection of the ischemic bowel and the formation of a proximal stoma.

Onset of mesenteric venous thrombosis can be acute or subacute based on the location of thrombosis in the splanchnic circulation. Patients often present with vague abdominal pain associated with nausea and vomiting. Physical examination findings include abdominal distention with mild to moderate tenderness and signs of dehydration. Findings on CT venous phase include diffuse bowel-wall thickening and thrombus within the splanchnic system. IV therapeutic anticoagulation, broad-spectrum antibiotics, and correction of electrolyte abnormalities should be performed. Surgical intervention is not performed unless there is evidence of peritonitis and/or bowel perforation. If there is evidence of bowel compromise, an exploratory laparotomy should be

performed with resection of compromised bowel. Second-look laparotomy after 24–48 h should be attempted as anticoagulation can help prevent resection of viable bowel. Hypercoagulability testing should be performed, and if underlying inherited disorders are diagnosed, lifelong anticoagulation is recommended.

A

We thank Cory Sandore for providing the illustration for this chapter. Susan Gearhart contributed to this chapter in the 18th edition, Rizwan Ahmed contributed to the 19th edition and Satinderjit Locham to the 20th edition.

■ FURTHER READING

- D QW et al: Risk factors for postoperative acute mesenteric ischemia among adult patients undergoing cardiac surgery: A systematic review and meta-analysis. *J Crit Care* 42:294, 2017.
S G et al: What is the best revascularization strategy for acute occlusive arterial mesenteric ischemia: Systematic review and meta-analysis. *Cardiovasc Interv Radiol* 41:27, 2018.
S MJ: Acute mesenteric ischemia. *Surg Clin North Am* 94:165, 2014.

330

Acute Intestinal Obstruction

Danny O. Jacobs



■ EPIDEMIOLOGY

Morbidity and mortality from acute intestinal obstruction have been decreasing over the past several decades. Nevertheless, the diagnosis can still be challenging, and the type of complications that patients suffer has not changed significantly. The extent of mechanical obstruction is typically described as partial, high grade, or complete—generally correlating with the risk of complications and the urgency with which the underlying disease process must be addressed. Obstruction is also commonly described as being either “simple” or, alternatively, “strangulated” if vascular insufficiency and intestinal ischemia are evident.

Acute intestinal obstruction occurs either *mechanically* from blockage or from intestinal dysmotility when there is no blockage. In the latter instance, the abnormality is described as being *functional*. Mechanical bowel obstruction may be caused by extrinsic processes, intrinsic abnormalities of the bowel wall, or intraluminal abnormalities (**Table 330-1**). Within each of these broad categories are many diseases that can impede intestinal propulsion. Intrinsic diseases that can cause intestinal obstruction are usually congenital, inflammatory, neoplastic, or traumatic in origin, although intussusception and radiation injury can also be etiologic.

Acute intestinal obstruction accounts for ~1–3% of all hospitalizations and a quarter of all urgent or emergent general surgery admissions. Approximately 80% of cases involve the small bowel, and about one-third of these patients show evidence of significant ischemia. The mortality rate for patients with strangulation who are operated on within 24–30 h of the onset of symptoms is ~8% but triples shortly thereafter.

Extrinsic diseases most commonly cause mechanical obstruction of the small intestine. In the United States and Europe, almost all cases are caused by postoperative adhesions, carcinomatosis, or herniation of the anterior abdominal wall. Carcinomatosis most often originates from the ovary, pancreas, stomach, or colon, although rarely, metastasis from distant organs like the breast and skin can occur. Adhesions are responsible for the majority of cases of early postoperative obstruction that require intervention. It is important to note many patients who are successfully treated for adhesive small-bowel obstruction will experience recurrence. Approximately 20% of patients who were treated

TABLE 330-1 Most Common Causes of Acute Intestinal Obstruction

Extrinsic Disease

Adhesions (especially due to previous abdominal surgery), internal or external hernias, neoplasms (including carcinomatosis and extraintestinal malignancies, mostly commonly ovarian), endometriosis or intraperitoneal abscesses, and idiopathic sclerosis

Intrinsic Disease

Congenital (e.g., malrotation, atresia, stenosis, intestinal duplication, cyst formation, and congenital bands—the latter rarely in adults)

Inflammation (e.g., inflammatory bowel disease, especially Crohn's disease, but also diverticulitis, radiation, tuberculosis, lymphogranuloma venereum, and schistosomiasis)

Neoplasia (note: primary small-bowel cancer is rare; obstructive colon cancer may mimic small-bowel obstruction if the ileocecal valve is incompetent)

Traumatic (e.g., hematoma formation, anastomotic strictures)

Other, including intussusception (where the lead point is typically a polyp or tumor in adults), volvulus, obstruction of duodenum by superior mesenteric artery, radiation or ischemic injury, and aganglionosis, which is Hirschsprung's disease

Intraluminal Abnormalities

Bezoars, feces, foreign bodies including inspissated barium, gallstones (entering the lumen via a cholecystoenteric fistula), enteroliths

conservatively and between 5 and 30% of patients who were managed operatively will require readmission within 10 years.

Open operations of the lower abdomen, including appendectomy and colorectal and gynecologic procedures, are especially likely to create adhesions that can cause bowel obstruction (Table 330-2). The risk of internal herniation is increased by abdominal procedures such as laparoscopic or open Roux-en-Y gastric bypass. Although laparoscopic procedures may generate fewer postoperative adhesions compared with open surgery, the risk of obstructive adhesion formation is not eliminated.

Volvulus, which occurs when bowel twists on its mesenteric axis, can cause partial or complete obstruction and vascular insufficiency. The sigmoid colon is most commonly affected, accounting for approximately two-thirds of all cases of volvulus and 4% of all cases of large-bowel obstruction. The cecum and terminal ileum can also volvulize, or the cecum alone may be involved as a cecal bascule. Risk factors include institutionalization, the presence of neuropsychiatric conditions requiring psychotropic medication, chronic constipation, and aging; patients typically present in their seventies or eighties.

Colonic volvulus is more common in Eastern Europe, Russia, and Africa than it is in the United States. It is rare for adhesions or hernias to obstruct the colon. Cancer of the descending colon and rectum is responsible for approximately two-thirds of all cases, followed by diverticulitis and volvulus.

Functional obstruction, also known as *ileus* and *pseudo-obstruction*, is present when dysmotility prevents intestinal contents from being propelled distally and no mechanical blockage exists. Ileus that occurs after intraabdominal surgery is the most commonly identified form of functional bowel obstruction, but there are many other causes (Table 330-3). Although postoperative ileus is most often transient, it

TABLE 330-3 Most Common Causes of Ileus (Functional or Pseudo-Obstruction of the Intestine)

Intraabdominal procedures, lumbar spinal injuries, or surgical procedures on the lumbar spine and pelvis

Metabolic or electrolyte abnormalities, especially hypokalemia and hypomagnesemia, but also hyponatremia, uremia, and severe hyperglycemia

Drugs such as opiates, antihistamines, and some psychotropic (e.g., haloperidol, tricyclic antidepressants) and anticholinergic agents

Intestinal ischemia

Intraabdominal or retroperitoneal inflammation or hemorrhage

Lower lobe pneumonias

Intraoperative radiation (likely due to muscle damage)

Systemic sepsis

Hyperparathyroidism

Pseudo-obstruction (Ogilvie's syndrome)

Ileus secondary to hereditary or acquired visceral myopathies and neuropathies that disrupt myocellular neural coordination

Some collagen vascular diseases such as lupus erythematosus or scleroderma

is often the most common reason why hospital discharge is delayed. Pseudo-obstruction of the colon, also known as Ogilvie's syndrome, is a relatively rare disease. Some patients with Ogilvie's syndrome have colonic dysmotility due to abnormalities of their autonomic nervous system that may be inherited.

■ PATHOPHYSIOLOGY

The manifestations of acute intestinal obstruction depend on the nature of the underlying disease process, its location, and changes in blood flow (Fig. 330-1). Increased intestinal contractility, which occurs proximally and distal to the obstruction, is a characteristic response. Subsequently, intestinal peristalsis slows as the intestine or stomach proximal to the point of obstruction dilates and fills with gastrointestinal secretions and swallowed air. Although swallowed air is the primary contributor to intestinal distension, intraluminal air may also accumulate from fermentation, local carbon dioxide production, and altered gaseous diffusion.

Intraluminal dilation also increases intraluminal pressure. When luminal pressure exceeds venous pressure, venous and lymphatic drainage is impeded. Edema ensues, and the bowel wall proximal to the site of blockage may become hypoxic. Epithelial necrosis can be identified within 12 h of obstruction. Ultimately, arterial blood supply may become so compromised that full-thickness ischemia, necrosis, and perforation result. Stasis increases the bacteria counts within the jejunum and ileum. Bacteria, such as *Escherichia coli*, *Streptococcus faecalis*, and *Klebsiella*, and other pathogens may be recovered from intestinal cultures, mesenteric lymph nodes, the bloodstream, and other sites.

Other manifestations depend on the degree of hypovolemia, the patient's metabolic response, and the presence or absence of associated intestinal ischemia. Inflammatory edema eventually increases the production of reactive oxygen species and activates neutrophils and macrophages, which accumulate within the bowel wall. Their accumulation, along with changes in innate immunity, disrupts secretory and neuromotor processes. Dehydration is caused by loss of the normal intestinal absorptive capacity as well as fluid accumulation in the gastric or intestinal wall and intraperitoneally.

Anorexia and emesis tend to exacerbate intravascular volume depletion. In the worst-case scenario that is most commonly identified after high-grade distal obstruction, emesis leads to losses of gastric potassium, hydrogen, and chloride, while dehydration stimulates proximal renal tubule bicarbonate reabsorption. Intraperitoneal fluid accumulation, especially in patients with severe distal bowel obstruction, may increase intraabdominal pressure enough to elevate the diaphragm, inhibit respiration, and impede systemic venous return and promote vascular instability. Severe hemodynamic compromise may elicit a systemic inflammatory response and generalized microvascular leakage.

TABLE 330-2 Acute Small-Intestinal and Colonic Obstruction Incidences

CAUSE	INCIDENCE
Postoperative adhesions	>50%
Neoplasms	~20%
Hernias (especially ventral or internal types, where the risk of strangulation is increased)	~10%
Inflammatory bowel disease, other inflammation (obstruction may resolve if acute inflammation and edema subside)	~5%
Intussusception, volvulus, other miscellaneous diseases	<15%

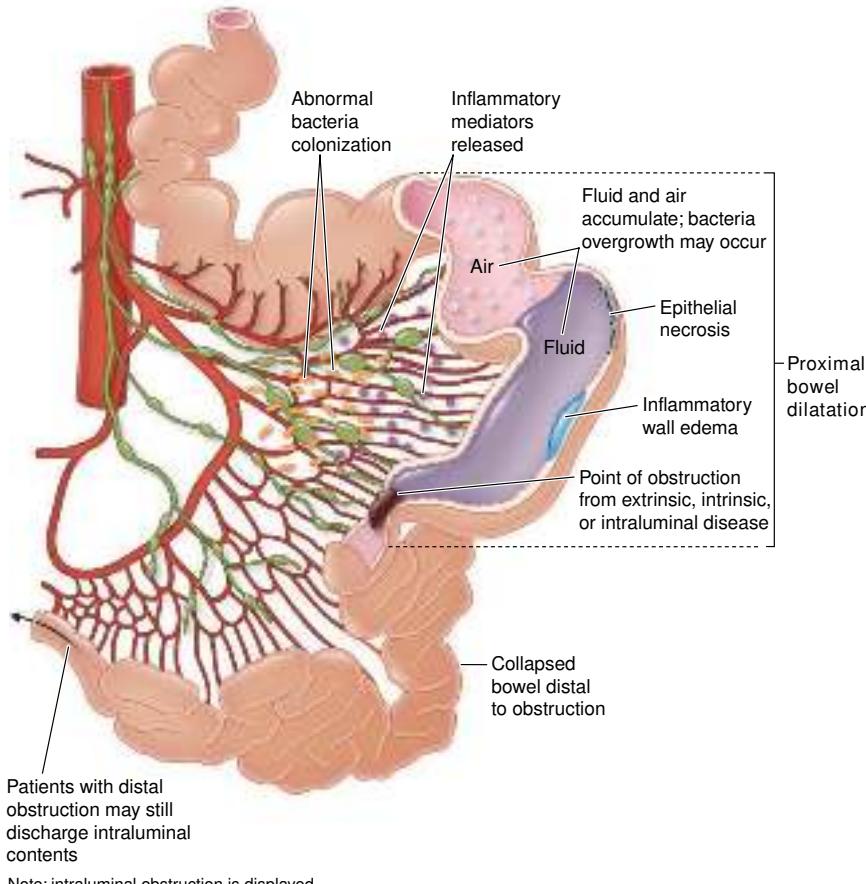


FIGURE 330-1 Pathophysiologic changes of small-bowel obstruction.

Closed-loop obstruction results when the proximal and distal openings of a given bowel segment are both occluded, for example, due to volvulus or a hernia. It is the most common precursor for strangulation, but not every closed loop strangulates. The risk of vascular insufficiency, systemic inflammation, hemodynamic compromise, and irreversible intestinal ischemia is much greater in patients with closed-loop obstruction. Pathologic changes may occur more rapidly, and emergency intervention is indicated. Irreversible bowel ischemia may progress to transmural necrosis even if obstruction is relieved. It is also important to remember that patients with high-grade distal colonic obstruction who have competent ileocecal valves may present with closed-loop obstruction. In the latter instance, the cecum may progressively dilate such that ischemic necrosis results in perforation especially when the cecal diameter exceeds 10–12 cm, as informed by Laplace's law. Patients with distal colonic obstruction whose ileocecal valves are incompetent tend to present later in the course of disease and mimic patients with distal small-bowel obstruction.

HISTORY AND PHYSICAL FINDINGS

Even though the presenting signs and symptoms can be misleading, many patients with acute obstruction can be accurately diagnosed after a thorough history and physical examination is performed. However, small-bowel obstruction with strangulation can be especially difficult to diagnosis promptly. Early recognition allows earlier treatment that decreases the risk of progression or other excess morbidity.

The cardinal signs are colicky abdominal pain, abdominal distension, emesis, and obstipation. More intraluminal fluid accumulates in patients with distal obstruction, which typically leads to greater distension, more discomfort, and delayed emesis. This emesis is feculent

when there is bacterial overgrowth. Patients with more proximal obstruction commonly present with less abdominal distension but more pronounced vomiting. Elements of the history that might be helpful include any prior history of surgery, including herniorrhaphy, as well as any history of cancer or inflammatory bowel disease.

Most patients, even those with simple obstruction, appear to be critically ill. Many may be oliguric, hypotensive, and tachycardic because of severe intravascular volume depletion. Fever is worrisome for strangulation or systemic inflammation. Bowel sounds and bowel functional activity are notoriously difficult to interpret. Classically, many patients with early small-bowel obstruction will have high-pitched, "musical" tinkling bowel sounds and peristaltic "rushes" known as borborygmi. Later in the course of disease, the bowel sounds may be absent or hypoactive as peristaltic activity decreases. This is in contrast to the common findings in patients with ileus or pseudo-obstruction where bowel sounds are typically absent or hypoactive from the beginning. Lastly, patients with partial blockage may continue to pass flatus and stool, and those with complete blockage may evacuate bowel contents present downstream beyond their obstruction.

All surgical incisions should be examined, and the presence of a tender abdominal or groin mass strongly suggests that an incarcerated hernia may be the cause of obstruction. The presence of tenderness should increase the concern about the presence of complications such as ischemia, necrosis, or peritonitis. Severe pain with localization or signs of peritoneal irritation is suspicious for strangulated or closed-loop obstruction. It is important to remember that the discomfort may be out of proportion to physical findings mimicking the complaints of patients with acute mesenteric ischemia. Patients with colonic volvulus present with the classic manifestations of closed-loop obstruction:

severe abdominal pain, vomiting, and obstipation. Asymmetrical abdominal distension and a tympanic mass may be evident.

Patients with ileus or pseudo-obstruction may have signs and symptoms similar to those of bowel obstruction. Although abdominal distension is present, colicky abdominal pain is typically absent, and patients may not have nausea or emesis. Ongoing, regular discharge of stool or flatus can sometimes help distinguish patients with ileus from those with complete mechanical bowel obstruction.

■ LABORATORY AND IMAGING STUDIES

Laboratory testing should include a complete blood count and serum electrolyte and creatinine measurements. Serial assessments are often useful. Mild hemoconcentration and slight elevation of the white blood cell count commonly occur after simple bowel obstruction. Emesis and dehydration may cause hypokalemia, hypochloremia, elevated blood urea nitrogen-to-creatinine ratios, and metabolic alkalosis. Patients may be hyponatremic on admission because many have attempted to rehydrate themselves with hypotonic fluids. The presence of guaiac-positive stools and iron-deficiency anemia are strongly suggestive of malignancy.

Higher white blood cell counts with the presence of immature forms or the presence of metabolic acidosis are worrisome for severe volume depletion or ischemic necrosis and sepsis. Presently, there are no laboratory tests that are especially useful for identifying the presence of simple or strangulated obstruction, although increases in serum β -lactate, creatine kinase BB isoenzymes, or intestinal fatty acid binding protein levels may be suggestive of the latter.

Recommendations for diagnostic imaging continue to evolve. In all cases, the key is not to delay operative intervention unnecessarily when the patient's signs or symptoms strongly suggest that high-grade or complete obstruction or bowel compromise is present. Abdominal radiography, which must include upright or cross-table lateral views, can be completed quickly and may indicate the need for emergency surgical intervention in patients who are not in the immediate post-operative period. A "staircasing" pattern of dilated air and fluid-filled small-bowel loops >2.5 cm in diameter with little or no air seen in the colon are classical findings in patients with small-bowel obstruction, although findings may be equivocal in some patients with documented disease. Little bowel gas appears in patients with proximal bowel obstruction or in patients whose intestinal lumens are filled with fluid. Upright plain films of the abdomen of patients with large-bowel obstruction typically show colon dilatation. Small-bowel air-fluid levels may not be obvious if the ileocecal valve is incompetent. Although it can be difficult to distinguish from ileus, small-bowel obstruction is more likely when air-fluid levels are seen without significant colonic distension. Free air suggests that perforation has occurred in patients who have not recently undergone surgical procedures. A gas-filled, "coffee bean"-shaped dilated shadow may be seen in patients with volvulus.

More sophisticated imaging, which may be unnecessarily time consuming and expensive, can nevertheless be beneficial when the diagnosis is unclear. Computed tomography (CT) is the most commonly used imaging modality. Its sensitivity for detecting bowel obstruction is ~95% (78–100%) in patients with high-grade obstruction, with a specificity of 96% and an accuracy of ~95%. Its accuracy in diagnosing closed-loop obstruction is much lower (60%). CT may also provide useful information regarding location or to identify particular circumstances where surgical intervention is needed urgently. Patients who have evidence of contrast appearing within the cecum within 4–24 h of oral administration of water-soluble contrast can be expected to improve with high sensitivity and specificity (~95% each). For example, contrast studies may demonstrate a "bird's beak," a "c-loop," or "whorl" deformity on CT imaging at the site where twisting obstructs the lumen when a colonic volvulus is present. Although abdominal radiography is usually the initial examination, unlike CT imaging, it may not accurately distinguish obstruction from other causes of colonic dysmotility. Examples of some CT images are reproduced in Fig. 330-2.

Ultrasonographic evaluations are especially difficult to interpret but may be sensitive and appropriate studies to evaluate patients who are

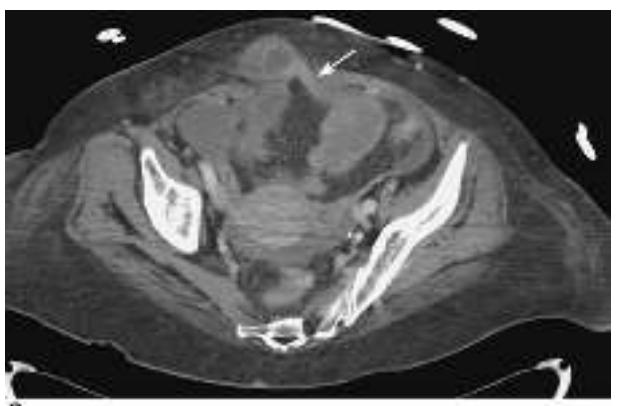
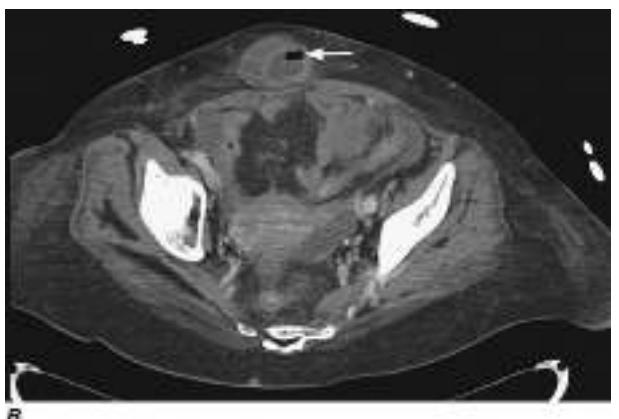


FIGURE 330-2 Computed tomography with oral and intravenous contrast demonstrating (A) evidence of small-bowel dilatation with air-fluid levels consistent with a small-bowel obstruction; (B) a partial small-bowel obstruction from an incarcerated ventral hernia (arrow); and (C) decompressed bowel seen distal to the hernia (arrow). (Reproduced with permission from D Longo et al: *Harrison's Principles of Internal Medicine*, 18th ed. New York: McGraw-Hill; 2012.)

pregnant or for whom x-ray exposure is otherwise contraindicated or inappropriate.

CT imaging with enteral and IV contrast can also identify ischemia. Altered bowel wall enhancement is the most specific early finding, but its sensitivity is low. Mesenteric venous gas, pneumoperitoneum, and pneumatosis intestinalis are late findings indicating the presence of bowel necrosis. CT scanning after a water-soluble contrast enema may help distinguish ileus or pseudo-obstruction from distal large-bowel obstruction in patients who present with evidence of small-bowel and colonic distention. CT enteroclysis, though rarely performed, can

2512 accurately identify neoplasia as a cause of bowel obstruction. Contrast enemas or colonoscopies are almost always needed to identify causes of acute colonic obstruction.

Barium studies are generally contraindicated in patients with firm evidence of complete or high-grade bowel obstruction, especially when they present acutely. Barium should never be given orally to a patient with possible obstruction until that diagnosis has been excluded. In every other case, such investigations should only be performed in exceptional circumstances and with great caution because patients with significant obstruction may develop barium concretions as an additional source of blockage and some who would have otherwise recovered will require operative intervention. Barium opacification also renders cross-sectional imaging studies or angiography uninterpretable.

TREATMENT

Acute Intestinal Obstruction

An improved understanding of the pathophysiology of bowel obstruction and the importance of fluid resuscitation, electrolyte repletion, intestinal decompression, and the selected use of antibiotics has likely contributed to a reduction in mortality from acute bowel obstruction. Every patient should be stabilized as quickly as possible. Nasogastric tube suction decompresses the stomach, minimizes further distention from swallowed air, improves patient comfort, and reduces the risk of aspiration. Urine output should be assessed using a Foley catheter. In some cases, for example, in patients with cardiac disease, central venous pressures should be monitored. The use of antibiotics is controversial, although prophylactic administration may be warranted if operation is anticipated. Complete bowel obstruction is an indication for intervention. Stenting may be possible and warranted for some patients with high-grade obstruction due to unresectable stage IV malignancy. Stenting may also allow elective mechanical bowel preparation before surgery is undertaken. Because treatment options are so variable, it is helpful to make as precise a diagnosis as possible preoperatively.

ILEUS

Patients with ileus are treated supportively with intravenous fluids and nasogastric decompression while any underlying pathology is treated. Pharmacologic therapy is not yet proven to be efficacious or cost-effective. However, peripherally active μ -opioid receptor antagonists (e.g., alvimopan and methylnaltrexone) may accelerate gastrointestinal recovery in some patients who have undergone abdominal surgery.

COLONIC PSEUDO-OBSTRUCTION (OGILVIES DISEASE)

Neostigmine is an acetylcholinesterase inhibitor that increases cholinergic (parasympathetic) activity, which can stimulate colonic motility. Some studies have shown it to be moderately effective in alleviating acute colonic pseudo-obstruction. It is the most common therapeutic approach and can be used once it is certain that there is no mechanical obstruction. Cardiac monitoring is required, and atropine should be immediately available. Intravenous administration induces defecation and flatus within 10 min in the majority of patients who will respond. Sympathetic blockade by epidural anesthesia can successfully ameliorate pseudo-obstruction in some patients.

COLVOLVULUS

Patients with sigmoid volvulus can often be decompressed using a flexible tube inserted through a rigid proctoscope or using a flexible sigmoidoscope. Successful decompression results in sudden release of gas and fluid with evidence of decreased abdominal distension and allows definitive correction to be scheduled electively. Cecal volvulus most often requires laparotomy or laparoscopic correction.

INTRAOPERATIVE STRATEGIES

Approximately 60–80% of selected patients with mechanical bowel obstruction can be successfully treated conservatively. Indeed, most cases of radiation-induced obstruction should also be managed

nonoperatively if possible. In most circumstances, early consultation with a surgeon is prudent when there is concern about strangulation obstruction or other abnormality that needs to be addressed urgently. Deterioration signifies a need for intervention. At this time, the decision as to whether the patient can continue to be treated nonoperatively can only be based on clinical judgment, although, as described earlier, imaging studies can sometimes be helpful. The frequency of major complications after operation ranges from 12 to 47%, with greater risk being attributed to resection therapies and the patient's overall health. Risk is increased for patients with American Society of Anesthesiologists (ASA) physical status of class III or higher.

At operation, dilation proximal to the site of blockage with distal collapse is a defining feature of bowel obstruction. Intraoperative strategies depend on the underlying problem and range from lysis of adhesions to resection with or without diverting ostomy to primary resection with anastomosis. Resection is warranted when there is concern about the bowel's viability after the obstructive process is relieved. Laparoscopic approaches can be useful for patients with early obstruction when extensive adhesions are not expected to be present. Some patients with high-grade obstruction secondary to malignant disease that is not amendable to resection will benefit from bypass procedures.

ADULT INTUSSUSCEPTION AND GALLSTONE ILEUS

Primary resection is prudent. Careful manual reduction of any involved bowel may limit the amount of intestine that needs to be removed. A proximal ostomy may be required if unprepped colon is involved. The most common site of intestinal obstruction in patients with gallstone "ileus" is the ileum (60% of patients). The gallstone enters the intestinal tract most often via a cholecystoduodenal fistula. It can usually be removed by operative enterolithotomy. Addressing the gallbladder disease during urgent or emergent surgery is not recommended.

POSTOPERATIVE BOWEL OBSTRUCTION

Early postoperative mechanical bowel obstruction is that which occurs within the first 6 weeks of operation. Most are partial and can be expected to resolve spontaneously. It tends to respond and behave differently from classic mechanical bowel obstruction and may be very difficult to distinguish from postoperative ileus. A higher index of suspicion for a definitive site of obstruction is warranted for patients who undergo laparoscopic surgical procedures. Patients who first had ileus and then subsequently develop obstructive symptoms after an initial return of normal bowel function are more likely to have true postoperative small-bowel obstruction. The longer it takes for a patient's obstructive symptoms to resolve after hospitalization, the more likely the patient is to require surgical intervention.

A

The wisdom and expertise of Dr. William Silen are gratefully acknowledged.

FURTHER READING

- C F et al: Adhesive small bowel adhesions obstruction: Evolutions in diagnosis, management and prevention. World J Gastrointest Surg 27:222, 2016.
- F P et al: Surgery or stenting for colonic obstruction: A practice management guideline from the Eastern Association for the Surgery of Trauma. J Trauma Acute Care Surg 80:659, 2016.
- J T, T WM: Large-bowel obstruction in the adults: Classic radiographic and CT findings, etiology and mimics. Radiology 275:651, 2015.
- P EK, T WM: Review of small-bowel obstruction: The diagnosis and when to worry. Radiology 275:332, 2015.
- P H et al: Relative accuracy of emergency CT in adults with non-traumatic abdominal pain. Brit Inst Rad 89:20150416, 2016.
- T MR, L N: Adult small bowel obstruction. Acad Emerg Med 20:528, 2013.

331

Acute Appendicitis and Peritonitis

Danny O. Jacobs



ACUTE APPENDICITIS

■ INCIDENCE AND EPIDEMIOLOGY

Appendicitis occurs more frequently in westernized societies, but its incidence is decreasing for uncertain reasons. Nevertheless, acute appendicitis remains the most common emergency general surgical disease affecting the abdomen, with a rate of ~100 per 100,000 person-years in Europe and the Americas or ~11 cases per 10,000 people annually. Approximately 9% of men and 7% of women will experience an episode during their lifetime. Appendicitis occurs most commonly in 10- to 19-year-olds; however, the average age at diagnosis appears to be gradually increasing. Overall, 70% of patients are <30 years old, and most are men.

One of the more common complications and most important causes of excess morbidity and mortality is perforation, whether it is contained and localized or unconstrained within the peritoneal cavity. The incidence of perforated appendicitis (~20 cases per 100,000 person-years) may be increasing. The explanation for this trend is unknown. Approximately 20% of all patients will present with evidence of perforation, but the percentage risk is much higher in patients <5 or >65 years of age.

■ PATHOGENESIS OF APPENDICITIS AND APPENDICEAL PERFORATION

Appendicitis was first described in 1886 by Reginald Fitz. Its etiology is still not completely understood. Fecaliths, incompletely digested food residue, lymphoid hyperplasia, intraluminal scarring, tumors, bacteria, viruses, and inflammatory bowel disease have all been associated with inflammation of the appendix and appendicitis with potentially different outcomes depending on pathogenesis.

Although not proven, obstruction of the appendiceal lumen is believed to be an important step in the development of appendicitis—at least in some cases. Here, obstruction leads to bacterial overgrowth and luminal distension, with an increase in intraluminal pressure that can inhibit the flow of lymph and blood. Then, vascular thrombosis and ischemic necrosis with perforation of the distal appendix may occur. Therefore, perforation that occurs near the base of the appendix should raise concerns about another disease process. Most patients who will perforate do so before they are evaluated by surgeons.

Appendiceal fecaliths (or appendicoliths) are found in ~50% of patients with gangrenous appendicitis who perforate but are rarely identified in those who have simple disease. As mentioned earlier, the incidence of perforated, but not simple, appendicitis appears to be increasing. The rate of perforated and nonperforated appendicitis is correlated in men but not in women. Together these observations suggest that the underlying pathophysiologic processes are different and that simple appendicitis does not always progress to perforation. It appears that at least some cases of simple acute appendicitis may resolve spontaneously or with antibiotic therapy with limited risk of recurrent disease. The use of antibiotics to treat uncomplicated appendicitis continues to be studied intensively. Some data indicate that some patients who present with uncomplicated appendicitis based on computed tomography (CT) and who are treated with antibiotics alone will not experience a recurrence within a year. These findings highlight the importance of clinical decision-making and risk assessment when deciding and discussing treatment options with patients who presumably have simple disease, for example, deciding who is an appropriate candidate for nonoperative management and who is not. The latter is especially pertinent given the difficulty in assessing which patients might progress to perforation and which will not.

TABLE 331-1 Some Conditions That Mimic Appendicitis

Crohn's disease	Meckel's diverticulitis
Gallbladder disease	Mittelschmerz
Diverticulitis	Mesenteric adenitis
Ectopic pregnancy	Omental torsion
Endometriosis	Pancreatitis
Gastroenteritis or colitis	Lower lobe pneumonia
Gastric or duodenal ulceration	Pelvic inflammatory disease
Hepatitis	Ruptured ovarian cyst or other cystic disease of the ovaries
Kidney disease, including nephrolithiasis	Small-bowel obstruction
Liver abscess	Urinary tract infection

Increasingly it appears that there are two broad categories of patients with appendicitis—those with complicated disease like gangrene or perforation and those without. When perforation occurs, the resultant leak may be contained by the omentum or other surrounding tissues to form an abscess. Free perforation normally causes severe peritonitis. These patients may also develop infective suppurative thrombosis of the portal vein and its tributaries along with intrahepatic abscesses. The prognosis of the very unfortunate patients who develop this rare but dreaded complication is very poor.

■ CLINICAL MANIFESTATIONS

Improved diagnosis, supportive care, and surgical interventions are likely responsible for the remarkable decrease in the risk of mortality from simple appendicitis to currently <1%. Nevertheless, it is still important to identify patients who might have appendicitis as early as possible. Patients who have persistent symptoms that have not improved over 48 h may be more likely to perforate or develop other complications.

Appendicitis should be included in the differential diagnosis of abdominal pain for every patient in any age group unless it is certain that the organ has been previously removed (Table 331-1).

The appendix's anatomic location, which varies, may directly influence how the patient presents. Where the appendix can be "found" ranges from local differences in how the appendiceal body and tip lie relative to its attachment to the cecum (Figs. 331-1 and 331-2), to where the appendix is actually situated in the peritoneal cavity—for example, from its typical location in the right lower quadrant, to the pelvis, right flank, right upper quadrant (as may be observed during pregnancy), or even the left side of the abdomen for patients with malrotation or who have severely redundant colons.

Because the differential diagnosis of appendicitis is so extensive, deciding if a patient has appendicitis can be difficult (Table 331-2). Many patients may not present with the classically described history or

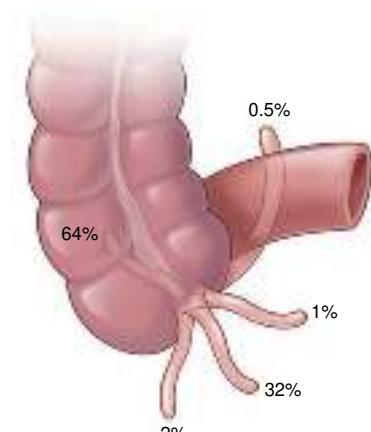


FIGURE 331-1 Regional anatomic variations of the appendix.



FIGURE 331-2 Locations of the appendix and cecum.

physical findings, and some may not have any abdominal discomfort early in the disease process. Soliciting an appropriate history requires detecting and evaluating symptoms that might suggest alternative diagnoses.

What is the classic history? Nonspecific complaints occur first. Patients may notice changes in bowel habits or malaise and vague, perhaps intermittent, crampy abdominal pain in the epigastric or periumbilical region. The pain subsequently migrates to the right lower quadrant over 12–24 h, where it is sharper and can be definitively localized as transmural inflammation when the appendix irritates the parietal peritoneum. Parietal peritoneal irritation may be associated with local muscle rigidity and stiffness. Patients with appendicitis will most often observe that their nausea, if present, followed the development of abdominal pain, which can help distinguish them from patients with gastroenteritis, for example, in whom nausea occurs first. Emesis, if present, also occurs after the onset of pain and is typically mild and scant. Thus, timing of the onset of symptoms and the characteristics of the patient's pain and any associated findings must be rigorously assessed. Anorexia is so common that the diagnosis of appendicitis should be questioned in its absence.

Arriving at the correct diagnosis is even more challenging when the appendix is not located in the right lower quadrant, in women of childbearing age, and in the very young or elderly. Because the differential diagnosis of appendicitis is so broad, often the key question

TABLE 331-2 Relative Frequency of Common Presenting Symptoms

SYMPTOMS	FREQUENCY
Abdominal pain	>95%
Anorexia	>70%
Constipation	4–16%
Diarrhea	4–16%
Fever	10–20%
Migration of pain to right lower quadrant	50–60%
Nausea	>65%
Vomiting	50–75%

TABLE 331-3 Relative Frequency of Some Presenting Signs

SIGNS	FREQUENCY
Abdominal tenderness	>95%
Right lower quadrant tenderness	>90%
Rebound tenderness	30–70%
Rectal tenderness	30–40%
Cervical motion tenderness	30%
Rigidity	~10%
Psoas sign	3–5%
Obturator sign	5–10%
Rovsing's sign	5%
Palpable mass	<5%

to answer expeditiously is whether the patient has appendicitis or some other condition that requires immediate operative intervention. A major concern is that the likelihood of a delay in diagnosis is greater if the appendix is unusually positioned. All patients should undergo a rectal examination. An inflamed appendix located behind the cecum or below the pelvic brim may prompt very little tenderness of the anterior abdominal wall.

Patients with pelvic appendicitis are more likely to present with dysuria, urinary frequency, diarrhea, or tenesmus. They may only experience pain in the suprapubic region on palpation or on rectal or pelvic examination. A pelvic examination in women is mandatory to rule out conditions affecting urogynecologic organs that can cause abdominal pain and mimic appendicitis such as pelvic inflammatory disease, ectopic pregnancy, and ovarian torsion. Interest in the ability of various clinical scoring systems to predict appendicitis or the need for imaging studies continues. However, none of the currently available decision tools yet appear to be able to circumvent or obviate the need for expert clinical opinion. The relative frequencies of some presenting signs are displayed in Table 331-3.

Patients with simple appendicitis normally only appear mildly ill with a pulse and temperature that are usually only slightly above normal. The provider should be concerned about other disease processes beside appendicitis or the presence of complications such as perforation, phlegmon, or abscess formation if the temperature is >38.3°C (~101°F) and if there are rigors.

Patients with appendicitis will be found to lie quite still to avoid peritoneal irritation caused by movement, and some will report discomfort caused by a bumpy car ride on the way to the hospital or clinic, coughing, sneezing, or other actions that replicate a Valsalva maneuver. The entire abdomen should be examined systematically starting in an area where the patient does not report discomfort if possible. Classically, maximal tenderness is identified where the appendix is most often located—in the right lower quadrant at or near McBurney's point, which is approximately one-third of the way along a line originating at the anterior iliac spine and running to the umbilicus. Gentle pressure in the left lower quadrant may elicit pain in the right lower quadrant if the appendix is located there. This is Rovsing's sign (Table 331-4). Evidence of parietal peritoneal irritation is often best elicited by gentle abdominal percussion, jiggling the patient's gurney or bed, or mildly bumping the feet.

TABLE 331-4 Classic Signs of Appendicitis in Patients with Abdominal Pain

MANEUVER	FINDINGS
Rovsing's sign	Palpating in the left lower quadrant causes pain in the right lower quadrant
Obturator sign	Internal rotation of the hip causes pain, suggesting the possibility of an inflamed appendix located in the pelvis
Iliopsoas sign	Extending the right hip causes pain along posterolateral back and hip, suggesting retrocecal appendicitis

Atypical presentation and pain patterns are common, especially in the very old or the very young. Diagnosing appendicitis in children can be especially challenging because they tend to respond so dramatically to stimulation and obtaining an accurate history may be difficult. In addition, it is important to remember that the smaller omentum found in children may be less likely to wall off an appendiceal perforation. Observing the child in a quiet surrounding may be helpful.

Signs and symptoms of appendicitis can be subtle in the elderly who may not react as vigorously to appendicitis as younger people. Pain, if noticed, may be minimal and have originated in the right lower quadrant or, otherwise, where the appendix is located. It may never have been noticed to be intermittent, or there may only be significant discomfort with deep palpation. Nausea, anorexia, and emesis may be the predominant complaints. The rare patient may even present with signs and symptoms of distal bowel obstruction secondary to appendiceal inflammation and phlegmon or abscess formation.

■ LABORATORY TESTING

Laboratory testing does not identify patients with appendicitis. The white blood cell count is only mildly to moderately elevated in ~70% of patients with simple appendicitis (with a leukocytosis of 10,000–18,000 cells/ μ L). A “left shift” toward immature polymorphonuclear leukocytes is present in >95% of cases. A sickle cell preparation may be prudent to obtain in those of African, Spanish, Mediterranean, or Indian ancestry. Serum amylase and lipase levels should be measured.

Urinalysis is indicated to help exclude genitourinary conditions that may mimic acute appendicitis, but a few red or white blood cells may be present as a nonspecific finding. However, an inflamed appendix that abuts the ureter or bladder may cause sterile pyuria or hematuria. Every woman of childbearing age should have a pregnancy test. Cervical cultures are indicated if pelvic inflammatory disease is suspected. Anemia and guaiac-positive stools should raise concern about the presence of other diseases or complications such as cancer.

■ IMAGING

Plain films of the abdomen are rarely helpful and so are not routinely obtained unless the clinician is worried about other conditions such as intestinal obstruction, perforated viscus, or ureterolithiasis. Less than 5% of patients will present with an opaque fecalith in the right lower quadrant. The presence of a fecalith is not diagnostic of appendicitis, although its presence in an appropriate location where the patient complains of pain is suggestive and is associated with a greater likelihood of complications.

The effectiveness of ultrasonography as a tool to diagnosis appendicitis is highly operator dependent. Even in very skilled hands, the appendix may not be visualized. Its overall sensitivity is 0.86, with a specificity of 0.81. Ultrasonography, especially intravaginal techniques, appears to be most useful for identifying pelvic pathology in women. Ultrasonographic findings suggesting the presence of appendicitis include wall thickening, an increased appendiceal diameter, and the presence of free fluid. Current practice in many institutions is to first perform ultrasonography and progress to other imaging studies only if the findings are equivocal.

The sensitivity and specificity of CT are at least 0.94 and 0.95, respectively. Thus, CT imaging, given its high negative predictive value, may be helpful if the diagnosis is in doubt, although studies performed early in the course of disease may not have any typical radiographic findings. In patients where the diagnosis is uncertain, delaying operation at the time of presentation to obtain CT does not appear to increase the risk of perforation. CT scanning is a superior method for assessing the severity of acute appendicitis in the absence of peritoneal findings indicative of perforation, abscess, or suspicion of an associated malignancy.

Suggestive findings on CT examination include dilatation >6 mm with wall thickening, a lumen that does not fill with enteric contrast, and fatty tissue stranding or air surrounding the appendix, which suggests inflammation (Figs. 331-3 and 331-4). The presence of luminal air or contrast is not consistent with a diagnosis of appendicitis. Furthermore, nonvisualization of the appendix is a nonspecific finding



FIGURE 331-3 Computed tomography with oral and intravenous contrast of acute appendicitis. There is thickening of the wall of the appendix and periappendiceal stranding (arrow).

that should not be used to rule out the presence of appendiceal or periappendiceal inflammation.

■ SPECIAL PATIENT POPULATIONS

Appendicitis is the most common extrauterine general surgical emergency observed during pregnancy. Early symptoms of appendicitis such as nausea and anorexia may be overlooked. Diagnosing appendicitis in pregnant patients may be especially difficult because as the uterus enlarges the appendix may be pushed higher along the right flank even to the right upper quadrant or because the gravid uterus may obscure typical physical findings. Ultrasonography may facilitate early diagnosis. A high index of suspicion is required because of the effects of unrecognized and untreated appendicitis on the fetus. For example, the fetal mortality rate is four times greater (from 5 to 20%) in patients with perforation.

Immunocompromised patients may present with only mild tenderness and may have many other disease processes in their differential diagnosis, including atypical infections from mycobacteria, *Cytomegalovirus*, or other fungi. Enterocolitis is a concern and may be present in patients who present with abdominal pain, fever, and neutropenia due to chemotherapy. CT imaging may be very helpful, although it is



FIGURE 331-4 Appendiceal fecalith (arrow).

TREATMENT

Acute Appendicitis

In the absence of contraindications, most patients who have strongly suggestive medical histories and physical examinations with supportive laboratory findings are candidates for appendectomy. In many instances, imaging studies are not required but are often obtained before surgical consultation is requested. Certainly, imaging and further study are appropriate in patients whose evaluations are suggestive but not convincing.

CT may accurately indicate the presence of appendicitis or other intraabdominal processes that warrant intervention. Whenever the diagnosis is uncertain, it is prudent to observe the patient and repeat the abdominal examination over 6–8 h. Any evidence of progression is an indication for operation. Narcotics can be given to patients with severe discomfort.

All patients should be fully prepared for surgery and have any fluid and electrolyte abnormalities corrected. Either laparoscopic or open appendectomy is a satisfactory choice for patients with uncomplicated appendicitis, although most procedures are now performed in a minimally invasive fashion to the patient's benefit in terms of recovery time and complications. Management of those who present with a mass representing a phlegmon or abscess can be more difficult. Such patients are best served by treatment with broad-spectrum antibiotics, drainage if there is an abscess >3 cm in diameter, and parenteral fluids and bowel rest if they appear to respond to conservative management. The appendix can then be more safely removed 6–12 weeks later when inflammation has diminished.

Laparoscopic appendectomy now accounts for the majority of all appendectomies performed in Western cultures and is associated with less postoperative pain, shorter lengths of stay, faster return to normal activity, and likely fewer superficial wound complications, although the risk of intraabdominal abscess formation may be higher.

A laparoscopic approach may also be useful when the exact diagnosis is uncertain. A laparoscopic approach may also facilitate exposure in those who are very obese. Absent complications, most patients can be discharged within 24–40 h of operation. The most common postoperative complications are fever and leukocytosis. Continuation of these findings beyond 5 days should raise concern for the presence of an intraabdominal abscess. The mortality rate for uncomplicated, nonperforated appendicitis is 0.1–0.5%, which approximates the overall risk of general anesthesia. The mortality rate for perforated appendicitis or other complicated disease is much higher, ranging from 3% overall to as high as 15% in the elderly.

ACUTE PERITONITIS

Acute peritonitis, or inflammation of the visceral and parietal peritoneum, is most often but not always infectious in origin, resulting from perforation of a hollow viscus. This is called *secondary peritonitis*, as opposed to *primary or spontaneous peritonitis*, when a specific intraabdominal source cannot be identified. In either instance, the inflammation can be localized or diffuse.

Etiology

Infective organisms may contaminate the peritoneal cavity after spillage from a hollow viscus, because of a penetrating wound of the abdominal wall, or because of the introduction of a foreign object like a peritoneal dialysis catheter or port that becomes infected. Secondary peritonitis most commonly results from perforation of the appendix, colonic diverticula, or the stomach and duodenum. It may also occur as a complication of bowel infarction or incarceration, cancer,

TABLE 331-5 Conditions Leading to Secondary Bacterial Peritonitis

Bowel perforation	Perforation or leakage of other organs
Appendicitis trauma (blunt or penetrating)	Biliary leakage (e.g., after liver biopsy)
Anastomotic leakage	Cholecystitis
Adhesion	Intraperitoneal bleeding
Diverticulitis	Pancreatitis
Iatrogenic (including endoscopic perforation)	Salpingitis
Ingested foreign body	Traumatic or other rupture of urinary bladder
Inflammation	Loss of peritoneal integrity
Intussusception	Intraperitoneal chemotherapy
Neoplasms	Iatrogenic (e.g., postoperative foreign body)
Obstruction	Perinephric abscess
Peptic ulcer disease	Peritoneal dialysis or other indwelling devices
Strangulated hernia	Trauma
Vascular (including ischemia or embolus)	

inflammatory bowel disease, and intestinal obstruction or volvulus. Conditions that may cause secondary bacterial peritonitis and their mechanisms are listed in Table 331-5. Over 90% of the cases of primary or spontaneous bacterial peritonitis occur in patients with ascites or hypoproteinemia (<1 g/L).

Aseptic peritonitis is most commonly caused by the abnormal presence of physiologic fluids such as gastric juice, bile, pancreatic enzymes, blood, or urine. It can also be caused by the effects of normally sterile foreign bodies such as surgical sponges or instruments. More rarely, it occurs as a complication of systemic diseases such as lupus erythematosus, porphyria, and familial Mediterranean fever. The chemical irritation caused by stomach acid and activated pancreatic enzymes is extreme, and secondary bacterial infection may occur.

CLINICAL FEATURES

The cardinal signs and symptoms of peritonitis are acute, typically severe, abdominal pain with tenderness and fever. How patients' complaints of pain are manifested depends on their overall physical health and whether the inflammation is diffuse or localized. Elderly and immunosuppressed patients may not respond as aggressively to the irritation. Diffuse, generalized peritonitis is most often recognized as diffuse abdominal tenderness with local guarding, rigidity, and other evidence of parietal peritoneal irritation. Physical findings may only be identified in a specific region of the abdomen if the intraperitoneal inflammatory process is limited or otherwise contained as may occur in patients with uncomplicated appendicitis or diverticulitis. Bowel sounds are usually absent to hypoactive.

Most patients present with tachycardia and signs of volume depletion with hypotension. Laboratory testing typically reveals a significant leukocytosis, and patients may be severely acidotic. Radiographic studies may show dilatation of the bowel and associated bowel wall edema. Free air or other evidence of leakage requires attention and could represent a surgical emergency. In stable patients in whom ascites is present, diagnostic paracentesis is indicated, where the fluid is tested for protein and lactate dehydrogenase and the cell count is measured.

THERAPY AND PROGNOSIS

Whereas mortality rates can be <10% for reasonably healthy patients with relatively uncomplicated, localized peritonitis, mortality rates >40% have been reported for the elderly or immunocompromised. Successful treatment depends on correcting any electrolyte abnormalities, restoration of fluid volume and stabilization of the cardiovascular system, appropriate antibiotic therapy, and surgical correction of any underlying abnormalities.

A

The wisdom and expertise of Dr. William Silen is gratefully acknowledged in this updated chapter on acute appendicitis and peritonitis.

FURTHER READING

- A RE: Short-term complications and long-term morbidity of laparoscopic and open appendicectomy in a national cohort. *Br J Surg* 101:1135, 2014.
- B MT et al: Changing epidemiology of acute appendicitis in the United States: Study period 1993–2008. *J Surg Res* 175:185, 2012.
- C CODA C : A randomized trial comparing antibiotics with appendectomy for appendicitis. *N Engl J Med* 383:1907, 2020.
- D S S et al: Diagnosis and treatment of acute appendicitis: 2020 update of the WSES Jerusalem guidelines. *World J Emerg Surg* 15:27, 2020.
- E FT et al: Time to appendectomy and risk of perforation in acute appendicitis. *JAMA Surg* 149:837, 2014.
- F DR: Acute appendicitis—appendectomy of the “antibiotics first” strategy. *N Engl J Med* 372:1937, 2015.
- G R et al: The Alvarado score for predicting acute appendicitis: A systematic review. *BMC Med* 9:139, 2011.
- H DA, DiSaverio S: Treatment of acute uncomplicated appendicitis. *N Engl J Med* 385:1116, 2021.
- I C et al: Amoxicillin plus clavulanic acid versus appendicectomy for treatment of acute uncomplicated appendicitis: An open-label, non-inferiority, randomised controlled trial. *Lancet* 377:1573, 2011.

men and ~1800 kcal/d for American women, although these estimates vary with body size and activity level. Formulas for roughly estimating REE are useful in assessing the energy needs of an individual whose weight is stable. Thus, for males, $REE = 900 + 10m$, and for females, $REE = 700 + 7m$, where m is mass in kilograms. The calculated REE is then adjusted for physical activity level by multiplying by 1.2 for sedentary, 1.4 for moderately active, or 1.8 for very active individuals. The final figure, the estimated energy requirement (EER), provides an approximation of total caloric needs in a state of energy balance for a person of a certain age, sex, weight, height, and physical activity level. For further discussion of energy balance in health and disease, see Chap. 334.

Protein Dietary protein consists of both essential and nonessential amino acids that are required for protein synthesis. The nine essential amino acids are histidine, isoleucine, leucine, lysine, methionine/cysteine, phenylalanine/tyrosine, threonine, tryptophan, and valine. Certain amino acids, such as alanine, can also be used for energy and gluconeogenesis. When energy intake is inadequate, protein intake must be increased, because ingested amino acids are diverted into pathways of glucose synthesis and oxidation. In extreme energy deprivation, protein-calorie malnutrition may ensue (Chap. 334).

For adults, the recommended dietary allowance (RDA) for protein is ~0.8 g/kg desirable body mass per day, assuming that energy needs are met and that the protein is of relatively high biologic value. Current recommendations for a healthy diet call for at least 10–14% of calories from protein. Most American diets provide at least those amounts. Biologic value tends to be highest for animal proteins, followed by proteins from legumes (beans), cereals (rice, wheat, corn), and roots. Combinations of plant proteins that complement one another in their essential amino acid profiles or combinations of animal and plant proteins can increase biologic value and lower total protein intakes necessary to meet requirements. In healthy people with adequate diets, the timing of protein intake over the course of the day has little effect.

Protein needs increase during growth, pregnancy, lactation, and rehabilitation after injury or undernutrition. Tolerance to dietary protein is decreased in renal insufficiency (with consequent uremia) and in liver failure. Usual protein intakes can precipitate encephalopathy in patients with cirrhosis of the liver.

Fat and Carbohydrate Fats are a concentrated source of energy and constitute, on average, 34% of calories in U.S. diets. However, for optimal health, fat intake should total no more than 30% of calories. Saturated fat and trans fat should be limited to <10% of calories and polyunsaturated fats to <10% of calories, with monounsaturated fats accounting for the remainder of fat intake. At least 45–55% of total calories should be derived from carbohydrates. The brain requires ~100 g of glucose per day for fuel; other tissues use ~50 g/d. Some tissues (e.g., brain and red blood cells) rely on glucose supplied either exogenously or from muscle proteolysis. Over time, during hypocaloric states, adaptations in carbohydrate needs are possible. Like fat (9 kcal/g), carbohydrate (4 kcal/g), and protein (4 kcal/g), alcohol (ethanol) provides energy (7 kcal/g). However, it is not a nutrient.

Water For adults, 1–1.5 mL of water per kilocalorie of energy expenditure is sufficient under usual conditions to allow for normal variations in physical activity, sweating, and solute load of the diet. Water losses include 50–100 mL/d in the feces; 500–1000 mL/d by evaporation or exhalation; and, depending on the renal solute load, ≥1000 mL/d in the urine. If external losses increase, intakes must increase accordingly to avoid underhydration. Fever increases water losses by ~200 mL/d per °C; diarrheal losses vary but may be as great as 5 L/d in severe diarrhea. Heavy sweating, vigorous exercise, and vomiting also increase water losses. When renal function is normal and solute intakes are adequate, the kidneys can adjust to increased water intake by excreting up to 18 L of excess water per day (Chap. 381). However, obligatory urine outputs can compromise hydration status when there is inadequate water intake or when losses increase in disease or kidney damage.

Infants have high requirements for water because of their large surface area to volume ratios, their inability to communicate their thirst, and the limited capacity of the immature kidney to handle high renal

Section 2 Nutrition

332

Nutrient Requirements and Dietary Assessment

Johanna T. Dwyer



Nutrients are substances that are not synthesized in sufficient amounts in the body and therefore must be supplied by the diet. Nutrient requirements for groups of healthy persons have been determined experimentally. The absence of essential nutrients leads to growth impairment, organ dysfunction, and failure to maintain nitrogen balance or adequate status of protein and other nutrients. For good health, we require energy-providing nutrients (protein, fat, and carbohydrate), vitamins, minerals, and water. Requirements for organic nutrients include 9 essential amino acids, several fatty acids, glucose, 4 fat-soluble vitamins, 10 water-soluble vitamins, dietary fiber, and choline. Several inorganic substances, including 4 minerals, 7 trace minerals, 3 electrolytes, and the ultratrace elements, must also be supplied by diet.

The amounts of essential nutrients required by individuals differ by their age and physiologic state. Conditionally essential nutrients are not required in the diet but must be supplied to certain individuals who do not synthesize them in adequate amounts, such as those with genetic defects; those with pathologies such as infection, disease, or trauma with nutritional implications; and developmentally immature infants. For example, inositol, taurine, arginine, and glutamine may be needed by premature infants. Many other organic and inorganic compounds that are present in foods and dietary supplements, including pesticides, lead, phytochemicals, zoochemicals, and microbial products, may also have health effects.

ESSENTIAL NUTRIENT REQUIREMENTS

Energy For weight to remain stable, energy intake must match energy output. The major components of energy output are resting energy expenditure (REE) and physical activity; minor components include the energy cost of metabolizing food (thermic effect of food, or specific dynamic action) and shivering thermogenesis (e.g., cold-induced thermogenesis). The average energy intake is ~2600 kcal/d for American

2518 solute loads. Increased water needs during pregnancy are ~30 mL/d. During lactation, milk production increases daily water requirements so that ~1000 mL of additional water is needed, or 1 mL for each milliliter of milk produced. Special attention must also be paid to the water needs of the elderly, who have reduced total-body water and blunted thirst sensation and are more likely to be taking medications such as diuretics.

Other Nutrients See Chap. 333 for detailed descriptions of vitamins and minerals.

■ DIETARY REFERENCE INTAKES AND RDAS

Fortunately, human life and well-being can be maintained within a fairly wide range with most nutrient intakes. However, the capacity for adaptation is not infinite—too much, as well as too little, intake of a nutrient can have adverse effects or alter the health benefits conferred by another nutrient. Therefore, benchmark recommendations regarding nutrient intakes have been developed to guide clinical practice. These quantitative estimates of nutrient intakes are collectively referred to in the United States and Canada as the *dietary reference intakes* (DRIs). The DRIs supplanted the RDAs—the single reference values used in the United States until the early 1990s. DRIs include an *estimated average requirement* (EAR) for nutrients as well as other reference values used for dietary planning: the RDA, the *adequate intake* (AI), the *chronic disease risk reduction intake* (CDRR), and the *tolerable upper level* (UL). The DRIs also include acceptable macronutrient distribution ranges (AMDRs) for protein, fat, and carbohydrate. The current DRIs for vitamins and elements are provided in Tables 332-1 and 332-2, respectively. Table 332-3 provides DRIs for water and macronutrients. EERs are discussed in Chap. 334 on energy balance in health and disease.

Estimated Average Requirement (EAR) When florid manifestations of the classic dietary-deficiency diseases such as rickets (deficiency of vitamin D and calcium), scurvy (deficiency of vitamin C), xerophthalmia (deficiency of vitamin A), and protein-calorie malnutrition were common, nutrient adequacy was inferred from the absence of their clinical deficiency signs. Later, biochemical and other changes were used that became evident long before the deficiency was clinically apparent. Consequently, criteria of adequacy are now based on biologic markers when they are available. Priority is given to sensitive biochemical, physiologic, or behavioral tests that reflect early changes in regulatory processes; maintenance of body stores of nutrients; or, if available, the amount of a nutrient that minimizes the risk of chronic degenerative disease. Current efforts focus on this last variable, but relevant markers often are not available, and the long time lags between intake and disease outcomes further complicate the picture.

The types of evidence and criteria used to establish nutrient requirements vary by nutrient, age, and physiologic group. The EAR is the amount of a nutrient estimated to be adequate for half of the healthy individuals of a specific age and sex. It is not an effective estimate of nutrient adequacy in individuals because it is a median requirement for a group; 50% of individuals in a group fall below the requirement and 50% fall above it. Thus, a person with a usual intake at the EAR has a 50% risk of inadequate intake. For these reasons, the other standards described below are more useful for clinical purposes.

Recommended Dietary Allowances The RDA, the nutrient intake goal for planning diets of individuals, is the average daily dietary intake level that meets the nutrient requirements of nearly all healthy persons of a specific sex, age, life stage, or physiologic condition (e.g., pregnancy or lactation). It is defined statistically as two standard deviations above the EAR to ensure that the needs of any given individual are met. An online tool, available at <https://www.nal.usda.gov/fnic/dri-calculator/>, allows health professionals to calculate individualized daily nutrient recommendations for dietary planning based on the DRIs. The RDAs are used to formulate food guides such as the U.S. Department of Agriculture (USDA) MyPlate Plan for individuals (<https://www.choosemyplate.gov/resources/MyPlatePlan>), to create food-exchange lists for therapeutic diet planning, and as a standard for describing the nutritional content of foods and nutrient-containing dietary supplements on labels.

The risk of dietary inadequacy increases as one's intake falls below the RDA. However, the RDA is an overly generous criterion for evaluating nutrient adequacy. For example, by definition, the RDA exceeds the actual requirements of all but ~2–3% of the population. Therefore, many people whose intake falls below the RDA are still getting enough of the nutrient. On food labels, the nutrient content in a food is stated by weight or as a percentage of the daily value (DV), a variant of the RDA used on the nutrition facts panel that, for an adult, represents the highest RDA for an adult consuming 2000 kcal.

Adequate Intake (AI) It is not possible to set an RDA for some nutrients that lack an established EAR. In this circumstance, the AI is based on observed or experimentally determined approximations of nutrient intakes in healthy people. In the DRIs, AIs rather than RDAs are proposed for nutrients consumed by infants (up to age 1 year) as well as for chromium, fluoride, manganese, sodium, potassium, pantothenic acid, biotin, choline, and water consumed by persons of all ages.

Tolerable Upper Levels (UL) Healthy individuals gain no established benefit from consuming nutrient levels above the RDA or AI. In fact, excessive nutrient intake can disturb body functions and cause acute, progressive, or permanent disabilities. The tolerable UL is the highest level of chronic (usually daily) nutrient intake that is unlikely to pose a risk of adverse health effects for most of the population. Data on the adverse effects of large amounts of many nutrients are unavailable or too limited to establish a UL. Therefore, the lack of a UL does *not* mean that the risk of adverse effects from high intake is nonexistent. Nutrient levels in commonly eaten foods rarely exceed the UL. However, very highly fortified foods and dietary supplements provide more concentrated amounts of nutrients per serving and thus pose a potential risk of toxicity. Dietary supplements are labeled with Supplement Facts that express the amount of nutrients present in absolute units or as the percentage of the DV provided per recommended serving size. Total nutrient intakes, including that in foods, supplements, and over-the-counter medications (e.g., antacids), should not exceed RDA levels.

Chronic Disease Risk Reduction Intake (CDRR) This is the level above which a reduction in intake is expected to lower chronic disease risk. For example, the sodium CDRR for adults is 2300 mg/d, and this is the lowest level of intake for which there is sufficiently strong evidence to characterize a CDRR. Three is no CDRR for potassium or other nutrients, but the AI for potassium has been reduced to 2500 mg/d from a higher prior level. At present, population recommendations for CDRR are not available for other nutrients.

Acceptable Macronutrient Distribution Ranges (AMDRs) AMDRs are not experimentally determined; rather, they are rough ranges for energy-providing macronutrient intakes (protein, carbohydrate, and fat) that the National Academy of Medicine's (formerly Institute of Medicine [IOM]) Food and Nutrition Board considers to be healthful. These ranges are 10–35% of calories for protein, 20–35% of calories for fat, and 45–65% of calories for carbohydrate. Alcohol, which also provides energy, is not a nutrient; therefore, no recommendations are provided.

■ FACTORS ALTERING NUTRIENT NEEDS

The DRIs are affected by age, sex, growth rate, pregnancy, lactation, physical activity level, concomitant diseases, drugs, and dietary composition. If requirements for nutrient sufficiency are close to intake levels indicating excess of a nutrient, dietary planning is difficult.

Physiologic Factors Growth, strenuous physical activity, pregnancy, and lactation all increase needs for energy and several essential nutrients. Energy needs rise during pregnancy due to fetal growth demands and increased energy required for milk production during lactation. Energy needs decrease with loss of lean body mass, the major determinant of REE. The energy needs of older persons, especially those aged >70 years, tend to be lower than those of younger persons because lean tissue, physical activity, and health often decline with age.

Dietary Composition Dietary composition affects the biologic availability and use of nutrients. For example, iron absorption may be

TABLE 332-1 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Vitamins

LIFE-STAGE GROUP	VITAMIN A (I g/d) ^a	VITAMIN C (mg/d)	VITAMIN D (I g/d) ^{b,c}	VITAMIN E (mg/d) ^d	VITAMIN K (I g/d)	THIAMIN (mg/d)	RIBOFLAVIN (mg/d)	NIACIN (mg/d) ^e	VITAMIN B ₆ (mg/d)	FOLATE (I g/d) ^f	VITAMIN B ₁₂ (I g/d)	PANTOTHENIC ACID (mg/d)	BIOTIN (I g/d)	CHOLINE (mg/d) ^g
Infants														
Birth to 6 mo	400 [*]	40 [*]	10	4 [*]	20 [*]	0.2 [*]	0.3 [*]	2 [*]	0.1 [*]	65 [*]	0.4 [*]	1.7 [*]	5 [*]	125 [*]
6–12 mo	500 [*]	50 [*]	10	5 [*]	25 [*]	0.3 [*]	0.4 [*]	4 [*]	0.3 [*]	80 [*]	0.5 [*]	1.8 [*]	6 [*]	150 [*]
Children														
1–3 y	300	15	15	6	30 [*]	0.5	0.5	6	0.5	150	0.9	2 [*]	8 [*]	200 [*]
4–8 y	400	25	15	7	55 [*]	0.6	0.6	8	0.6	200	1.2	3 [*]	12 [*]	250 [*]
Males														
9–13 y	600	45	15	11	60 [*]	0.9	0.9	12	1.0	300	1.8	4 [*]	20 [*]	375 [*]
14–18 y	900	75	15	15	75 [*]	1.2	1.3	16	1.3	400	24	5 [*]	25 [*]	550 [*]
19–30 y	900	90	15	15	120 [*]	1.2	1.3	16	1.3	400	24	5 [*]	30 [*]	550 [*]
31–50 y	900	90	15	15	120 [*]	1.2	1.3	16	1.3	400	24	5 [*]	30 [*]	550 [*]
51–70 y	900	90	15	15	120 [*]	1.2	1.3	16	1.7	400	24 ^h	5 [*]	30 [*]	550 [*]
>70 y	900	90	20	15	120 [*]	1.2	1.3	16	1.7	400	24 ^h	5 [*]	30 [*]	550 [*]
Females														
9–13 y	600	45	15	11	60 [*]	0.9	0.9	12	1.0	300	1.8	4 [*]	20 [*]	375 [*]
14–18 y	700	65	15	15	75 [*]	1.0	1.0	14	1.2	400	24	5 [*]	25 [*]	400 [*]
19–30 y	700	75	15	15	90 [*]	1.1	1.1	14	1.3	400	24	5 [*]	30 [*]	425 [*]
31–50 y	700	75	15	15	90 [*]	1.1	1.1	14	1.3	400	24	5 [*]	30 [*]	425 [*]
51–70 y	700	75	15	15	90 [*]	1.1	1.1	14	1.5	400	24 ^h	5 [*]	30 [*]	425 [*]
>70 y	700	75	20	15	90 [*]	1.1	1.1	14	1.5	400	24 ^h	5 [*]	30 [*]	425 [*]
Pregnant Women														
14–18 y	750	80	15	15	75 [*]	1.4	1.4	18	1.9	600	2.6	6 [*]	30 [*]	450 [*]
19–30 y	770	85	15	15	90 [*]	1.4	1.4	18	1.9	600	2.6	6 [*]	30 [*]	450 [*]
31–50 y	770	85	15	15	90 [*]	1.4	1.4	18	1.9	600	2.6	6 [*]	30 [*]	450 [*]
Lactating Women														
14–18 y	1200	115	15	19	75 [*]	1.4	1.6	17	2.0	500	2.8	7 [*]	35 [*]	550 [*]
19–30 y	1300	120	15	19	90 [*]	1.4	1.6	17	2.0	500	2.8	7 [*]	35 [*]	550 [*]
31–50 y	1300	120	15	19	90 [*]	1.4	1.6	17	2.0	500	2.8	7 [*]	35 [*]	550 [*]

Note: This table (taken from the DRI reports; see www.nap.edu) presents recommended dietary allowances (RDAs) in bold type and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data makes it impossible to specify with confidence the percentage of individuals covered by this intake.

^aAs retinol activity equivalents (RAEs). 1 RAE = 1 µg retinol, 12 µg β-carotene, 24 µg α-carotene, or 24 µg β-cryptoxanthin. The RAE for dietary provitamin A carotenoids is twofold greater than the retinol equivalent (RE), whereas the RAE for preformed vitamin A is the same as the RE. ^bAs cholecalciferol. 1 µg cholecalciferol = 40 IU vitamin D. ^cUnder the assumption of minimal sunlight. ^dAs α-tocopherol. α-Tocopherol includes RRα-α-tocopherol, the only form of α-tocopherol that occurs naturally in foods, and the 2R-stereoisomeric forms of α-tocopherol (RRR-, RSR-, RSS-, and RSSS-α-tocopherol) that occur in fortified foods and supplements. It does not include the 2S-stereoisomeric forms of α-tocopherol (SRR-, SSR-, SRS-, and SSS-α-tocopherol) also found in fortified foods and supplements. ^eAs niacin equivalents (NEs). 1 mg of niacin = 60 mg of tryptophan; 0–6 months = preformed niacin (not NE). ^fAs dietary folate equivalents (DFEs). 1 DFE = 1 µg food folate = 0.6 µg of folic acid from fortified food or as a supplement consumed with food = 0.5 µg of a supplement taken on an empty stomach. ^gAlthough AIs have been set for choline, there are few data to assess whether a dietary supply of choline is needed at all stages of the life cycle, and it may be that the choline requirement can be met by endogenous synthesis at some of these stages. ^hBecause 10–30% of older people may malabsorb food-bound B₁₂, it is advisable for those >50 years of age to meet their RDA mainly by consuming foods fortified with B₁₂ or a supplement containing B₁₂. In view of evidence linking inadequate folate intake with neural tube defects in the fetus, it is recommended that all women capable of becoming pregnant consume 400 µg of folate from supplements or fortified foods in addition to intake of food folate from a varied diet. ⁱIt is assumed that women will continue consuming 400 µg from supplements or fortified food until their pregnancy is confirmed and they enter prenatal care, which ordinarily occurs after the end of the periconceptional period—the critical time for formation of the neural tube.

Source: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. <https://doi.org/10.17226/25353>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

TABLE 332-2 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Elements

LIFE-STAGE GROUP	CALCIUM (mg/d)	CHROMIUM (l g/d)	COPPER (l g/d)	FLUORIDE (mg/d)	IODINE (l g/d)	IRON (mg/d)	MAGNESIUM (mg/d)	MANGANESE (mg/d)	MOLYBDENUM (l g/d)	PHOSPHORUS (mg/d)	SELENIUM (l g/d)	ZINC (mg/d)	POTASSIUM (g/d)	SODIUM (g/d)	CHLORIDE (g/d)
Infants															
Birth to 6 mo	200*	0.2*	200*	0.01*	110*	0.27*	30*	0.003*	2*	100*	15*	2*	0.4*	0.12*	0.18*
6–12 mo	260*	5.5*	220*	0.5*	130*	11	75*	0.6*	3*	275*	20*	3	0.7*	0.37*	0.57*
Children															
1–3 y	700	11*	340	0.7*	90	7	80	1.2*	17	460	20	3	3.0*	1.0*	1.5*
4–8 y	1000	15*	440	1*	90	10	130	1.5*	22	500	30	5	3.8*	1.2*	1.9*
Males															
9–13 y	1300	25*	700	2*	120	8	240	1.9*	34	1250	40	8	4.5*	1.5*	2.3*
14–18 y	1300	35*	890	3*	150	11	410	2.2*	43	1250	55	11	4.7*	1.5*	2.3*
19–30 y	1000	35*	900	4*	150	8	400	2.3*	45	700	55	11	4.7*	1.5*	2.3*
31–50 y	1000	35*	900	4*	150	8	420	2.3*	45	700	55	11	4.7*	1.5*	2.3*
51–70 y	1000	30*	900	4*	150	8	420	2.3*	45	700	55	11	4.7*	1.3*	2.0*
>70 y	1200	30*	900	4*	150	8	420	2.3*	45	700	55	11	4.7*	1.2*	1.8*
Females															
9–13 y	1300	21*	700	2*	120	8	240	1.6*	34	1250	40	8	4.5*	1.5*	2.3*
14–18 y	1300	24*	890	3*	150	15	360	1.6*	43	1250	55	9	4.7*	1.5*	2.3*
19–30 y	1000	25*	900	3*	150	18	310	1.8*	45	700	55	8	4.7*	1.5*	2.3*
31–50 y	1000	25*	900	3*	150	18	320	1.8*	45	700	55	8	4.7*	1.5*	2.3*
51–70 y	1200	20*	900	3*	150	8	320	1.8*	45	700	55	8	4.7*	1.3*	2.0*
>70 y	1200	20*	900	3*	150	8	320	1.8*	45	700	55	8	4.7*	1.2*	1.8*
Pregnant Women															
14–18 y	1300	29*	1000	3*	220	27	400	2.0*	50	1250	60	12	4.7*	1.5*	2.3*
19–30 y	1000	30*	1000	3*	220	27	350	2.0*	50	700	60	11	4.7*	1.5*	2.3*
31–50 y	1000	30*	1000	3*	220	27	360	2.0*	50	700	60	11	4.7*	1.5*	2.3*
Lactating Women															
14–18 y	1300	44*	1300	3*	290	10	360	2.6*	50	1250	70	13	5.1*	1.5*	2.3*
19–30 y	1000	45*	1300	3*	290	9	310	2.6*	50	700	70	12	5.1*	1.5*	2.3*
31–50 y	1000	45*	1300	3*	290	9	320	2.6*	50	700	70	12	5.1*	1.5*	2.3*

Note: This table (taken from the DRI reports; see www.nap.edu) presents recommended dietary allowances (RDAs) in bold type and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data makes it impossible to specify with confidence the percentage of individuals covered by this intake.

Sources: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. <https://doi.org/10.17226/25353>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

TABLE 332-3 Dietary Reference Intakes (DRIs): Recommended Dietary Allowances and Adequate Intakes for Total Water and Macronutrients

LIFE-STAGE GROUP	TOTAL WATER ^a (L/d)	CARBOHYDRATE (g/d)	TOTAL FIBER (g/d)	FAT (g/d)	LINOLEIC ACID (g/d)	-LINOLENIC ACID (g/d)	PROTEIN ^b (g/d)
Infants							
Birth to 6 mo	0.7 ^c	60 ^c	ND ^c	31 ^c	4.4 ^c	0.5 ^c	9.1 ^c
6–12 mo	0.8 ^c	95 ^c	ND	30 ^c	4.6 ^c	0.5 ^c	11.0 ^c
Children							
1–3 y	1.3 ^c	130	19 ^c	ND	7 ^c	0.7 ^c	13
4–8 y	1.7 ^c	130	25 ^c	ND	10 ^c	0.9 ^c	19
Males							
9–13 y	24 ^c	130	31 ^c	ND	12 ^c	1.2 ^c	34
14–18 y	33 ^c	130	38 ^c	ND	16 ^c	1.6 ^c	52
19–30 y	3.7 ^c	130	38 ^c	ND	17 ^c	1.6 ^c	56
31–50 y	3.7 ^c	130	38 ^c	ND	17 ^c	1.6 ^c	56
51–70 y	3.7 ^c	130	30 ^c	ND	14 ^c	1.6 ^c	56
>70 y	3.7 ^c	130	30 ^c	ND	14 ^c	1.6 ^c	56
Females							
9–13 y	2.1 ^c	130	26 ^c	ND	10 ^c	1.0 ^c	34
14–18 y	2.3 ^c	130	26 ^c	ND	11 ^c	1.1 ^c	46
19–30 y	2.7 ^c	130	25 ^c	ND	12 ^c	1.1 ^c	46
31–50 y	2.7 ^c	130	25 ^c	ND	12 ^c	1.1 ^c	46
51–70 y	2.7 ^c	130	21 ^c	ND	11 ^c	1.1 ^c	46
>70 y	2.7 ^c	130	21 ^c	ND	11 ^c	1.1 ^c	46
Pregnant Women							
14–18 y	3.0 ^c	175	28 ^c	ND	13 ^c	1.4 ^c	71
19–30 y	3.0 ^c	175	28 ^c	ND	13 ^c	1.4 ^c	71
31–50 y	3.0 ^c	175	28 ^c	ND	13 ^c	1.4 ^c	71
Lactating Women							
14–18	3.8 ^c	210	29 ^c	ND	13 ^c	1.3 ^c	71
19–30 y	3.8 ^c	210	29 ^c	ND	13 ^c	1.3 ^c	71
31–50 y	3.8 ^c	210	29 ^c	ND	13 ^c	1.3 ^c	71

Note: This table (taken from the DRI reports; see www.nap.edu) presents recommended dietary allowances (RDAs) in bold type and adequate intakes (AIs) in ordinary type followed by an asterisk (*). An RDA is the average daily dietary intake level sufficient to meet the nutrient requirements of nearly all healthy individuals (97–98%) in a group. The RDA is calculated from an estimated average requirement (EAR). If sufficient scientific evidence is not available to establish an EAR and thus to calculate an RDA, an AI is usually developed. For healthy breast-fed infants, an AI is the mean intake. The AI for other life-stage and sex-specific groups is believed to cover the needs of all healthy individuals in those groups, but lack of data or uncertainty in the data make it impossible to specify with confidence the percentage of individuals covered by this intake.

^aTotal water includes all water contained in food, beverages, and drinking water. ^bBased on grams of protein per kilogram of body weight for the reference body weight (e.g., for adults: 0.8 g/kg body weight for the reference body weight). ^cNot determined.

Source: National Academies of Sciences, Engineering, and Medicine. 2019. Dietary Reference Intakes for Sodium and Potassium. <https://doi.org/10.17226/25353>. Adapted and reproduced with permission from the National Academy of Sciences, Courtesy of the National Academies.

impaired by large amounts of calcium or lead; likewise, non-heme iron uptake may be impaired by a lack of ascorbic acid and amino acids in the meal. Bodily protein may be decreased when essential amino acids are not present in sufficient amounts—a rare scenario in U.S. diets. Animal foods, such as milk, eggs, and meat, have high biologic values, with most of the needed amino acids present in adequate amounts. Plant proteins in corn (maize), soy, rice, and wheat have lower biologic values and must be combined with other plant or animal proteins or fortified with the amino acids that are deficient to achieve optimal use by the body.

Route of Intake The RDAs apply only to oral intakes. When nutrients are administered parenterally, similar values can sometimes be used for amino acids, glucose (carbohydrate), fats, sodium, chloride, potassium, and most vitamins because their intestinal absorption rate is nearly 100%. However, the oral bioavailability of most mineral elements may be only half that obtained by parenteral administration. For some nutrients that are not readily stored in the body or that cannot be stored in large amounts, timing of administration may also be important. For example, amino acids cannot be used for protein synthesis if they are not supplied together; instead, they will be used for energy

production, although in healthy individuals eating adequate diets, the distribution of protein intake over the course of the day has little effect on health.

Disease Dietary deficiency diseases include protein-calorie malnutrition, iron-deficiency anemia, goiter (due to iodine deficiency), rickets and osteomalacia (vitamin D deficiency), xerophthalmia (vitamin A deficiency), megaloblastic anemia (vitamin B₁₂ or folic acid deficiency), scurvy (vitamin C/ascorbic acid deficiency), beriberi (thiamin deficiency), and pellagra (niacin and tryptophan deficiency) (**Chaps. 333 and 334**). Each deficiency disease is characterized by imbalances at the cellular level between the supply of nutrients or energy and the body's nutritional needs for growth, maintenance, and other functions. Imbalances and excesses in nutrient intakes are recognized as risk factors for certain chronic degenerative diseases, such as saturated fat and cholesterol in coronary artery disease; sodium in hypertension; obesity in hormone-dependent cancers (endometrial and breast); and ethanol in alcoholism. Diet is only one of many risk factors because the etiology and pathogenesis of these disorders are multifactorial. Osteoporosis, for example, is associated with calcium deficiency, sometimes secondary to vitamin D deficiency, as well as

2522 with environment-related risk factors (e.g., smoking, sedentary lifestyle), physiology (e.g., estrogen deficiency), genetic determinants (e.g., defects in collagen metabolism), and drug use (chronic steroids and aromatase inhibitors) (Chap. 411).

DIETARY ASSESSMENT

Nutrition assessment in clinical situations is an iterative process that involves (1) screening for malnutrition, (2) assessing the diet and other data to establish either the absence or the presence of malnutrition and its possible causes, (3) planning and implementing the most appropriate nutritional therapy, and (4) reassessing intakes to make sure that they have been consumed. Some disease states affect the bioavailability, requirements, use, or excretion of specific nutrients. In these circumstances, specific measurements of various nutrients or their biomarkers may be required to ensure adequate replacement (Chap. 333).

Most health care facilities have nutrition-screening processes in place for identifying possible malnutrition after hospital admission. Nutritional screening is required by The Joint Commission, which accredits and certifies health care organizations in the United States. However, no universally recognized or validated standards exist. The factors that are usually assessed include abnormal weight for height or body mass index (e.g., BMI <19 or >25); reported weight change (involuntary loss or gain of >5 kg in the past 6 months) (Chap. 47); diagnoses with known nutritional implications (e.g., metabolic disease, any disease affecting the gastrointestinal tract, alcoholism); present therapeutic dietary prescription; chronic poor appetite; presence of chewing and swallowing problems or major food intolerances; need for assistance with preparing or shopping for food, eating, or other aspects of self-care; and social isolation. The nutritional status of hospitalized patients should be reassessed periodically—at least once every week.

A more complete dietary assessment is indicated for patients who exhibit a high risk of or frank malnutrition on nutritional screening. The type of assessment varies with the clinical setting, the severity of the patient's illness, and the stability of the patient's condition.

Acute-Care Settings In acute-care settings, anorexia, various other diseases, test procedures, and medications can compromise dietary intake. Under such circumstances, the goal is to identify and avoid inadequate intake and to assure appropriate alimentation. Dietary assessment focuses on what patients are currently eating, whether or not they are able and willing to eat, and whether or not they experience any problems with eating. Dietary intake assessment is based on information from observed intakes; medical records; history; clinical examination; and anthropometric, biochemical, and functional status evaluations. The objective is to gather enough information to establish the likelihood of malnutrition due to poor dietary intake or other causes in order to assess whether nutritional therapy is indicated (Chap. 335).

Simple observations may suffice to suggest inadequate oral intake. These include dietitians' and nurses' notes; observation of a patient's frequent refusal to eat or the amount of food eaten on trays; the frequent performance of tests and procedures that are likely to cause meals to be skipped; adherence to nutritionally inadequate diet orders (e.g., clear liquids or full liquids) for more than a few days; the occurrence of fever, gastrointestinal distress, vomiting, diarrhea, or a comatose state; and the presence of diseases or use of treatments that involve any part of the alimentary tract. Acutely ill patients with diet-related diseases such as diabetes need assessment because an inappropriate diet may exacerbate these conditions and adversely affect other therapies. Abnormal biochemical values (serum albumin levels <35 g/L [$<3.5 \text{ mg/dL}$]; serum cholesterol levels <3.9 mmol/L [$<150 \text{ mg/dL}$]) are nonspecific but may indicate a need for further nutritional assessment.

Most therapeutic diets offered in hospitals are calculated to meet individual nutrient requirements and the RDA if they are eaten. Exceptions include clear liquids, some full-liquid diets, and test diets (such as those adhered to in preparation for gastrointestinal procedures), which are inadequate for several nutrients and should not be used, if possible, for >24 h. However, because as much as half of the food served to hospitalized patients is not eaten, it cannot be assumed that the intakes of hospitalized patients are adequate. Dietary assessment should compare

how much and what kinds of food the patient has consumed with the diet that has been provided. Major deviations in intakes of energy, protein, fluids, or other nutrients of special concern for the patient's illness should be noted and corrected, especially for long-staying patients.

Nutritional monitoring is especially important for patients who are very ill and who have extended lengths of hospital stay. Patients who are fed by enteral and parenteral routes also require special nutritional assessment and monitoring by physicians and/or dietitians with certification in nutritional support (Chap. 335).

Ambulatory Settings The aim of dietary assessment in the outpatient setting is to determine whether or not the patient's usual diet is a health risk in itself or if it contributes to existing chronic disease-related problems. Dietary assessment also provides the basis for planning a diet that fulfills therapeutic goals while ensuring patient adherence. The outpatient's dietary assessment should review the adequacy of present and usual food intakes, including vitamin and mineral supplements, oral nutritional supplements, medical foods, other dietary supplements, medications, and alcohol, because all of these may affect the patient's nutritional status. The assessment should focus on the dietary constituents that are most likely to be involved or compromised by a specific diagnosis as well as on any comorbidities that are present. More than 1 day's intake should be reviewed to provide a better representation of the usual diet, upon which personalized dietary recommendations can be based.

There are many ways to assess the adequacy of a patient's habitual diet. These include use of a food guide, a food-exchange list, a diet history, or a food-frequency questionnaire. A commonly used food guide for healthy persons is the USDA's Choose My Plate, which is useful as a rough guide for avoiding inadequate intakes of essential nutrients as well as likely excesses in the amounts of fat (especially saturated and trans fats), sodium, sugar, and alcohol consumed (Table 332-4). The Choose My Plate graphic emphasizes a balance between calories and

TABLE 332-4 Choose My Plate: A Guide to Individualized Dietary Planning

DIETARY FACTOR, UNIT OF MEASURE (ADVICE)	EXAMPLES OF STANDARD PORTION SIZES AT INDICATED ENERGY LEVEL		
	LOWER: 1600 kcal	MODERATE: 2200 kcal	HIGHER: 2800 kcal
Fruits, cups (Focus on fruits.)	1.5	2	2.5
Vegetables, cups (Vary vegetables.)	2	3	3.5
Grains, oz eq (Make at least half of grains whole.) ^a	5	7	10
Protein foods, oz eq (Go lean with protein.) ^b	5	6	7
Dairy, cups or oz ^c (Choose calcium-rich foods.)	3	3	3
"Empty" calories, kcal ^d	120	260	400
Sodium, mg	<2300 at all energy levels		
Physical activity, min	At least 150 min vigorous physical activity per week at all energy levels		

Note: Oils (formerly listed with portions of 5, 6, and 8 teaspoons for the lower, moderate, and higher energy levels, respectively) are no longer singled out in Choose My Plate, but rather are included in the empty calories/added sugar category with SOFAS (calories from solid fats and added sugars). The limit is the remaining number of calories in each food pattern above after intake of the recommended amounts of the nutrient-dense foods.

^aFor example, 1 serving equals 1 slice bread, 1 cup ready-to-eat cereal, or 0.5 cup cooked rice, pasta, or cooked cereal. ^bFor example, 1 serving equals 1 oz lean meat, poultry, or fish; 1 egg; 1 tablespoon peanut butter; 0.25 cup cooked dry beans; or 0.5 oz nuts or seeds. ^cFor example, 1 serving equals 1 cup milk or yogurt, 1.5 oz natural cheese, or 2 oz processed cheese. ^dFormerly called "discretionary calorie allowance." Portions are calculated as the number of calories remaining after all of the above allotments are accounted for.

Abbreviation: oz eq, ounce equivalent.

Source: Data from U.S. Department of Agriculture (<http://www.Choosemyplate.gov>).

nutritional needs, encouraging increased intake of fruits and vegetables, whole grains, and low-fat milk in conjunction with reduced intake of sodium and high-calorie sugary drinks. The Web version of the guide provides a calculator that tailors the number of servings suggested for healthy patients of different weights, sexes, ages, and life-cycle stages to help them to meet their needs while avoiding excess (<https://www.myplate.gov/myplate-plan> and www.ChooseMyPlate.gov). Patients who follow ethnic or unusual dietary patterns may need extra instruction on how foods should be categorized and on the appropriate portion sizes that constitute a serving. The process of reviewing the guide with patients helps them transition to healthier dietary patterns and identifies food groups eaten in excess of recommendations or in insufficient quantities. For persons on therapeutic diets, assessment against food-exchange lists may be useful. These include, for example, American Diabetes Association food-exchange lists for diabetes and the Academy of Nutrition and Dietetics food-exchange lists for renal disease.

NUTRITIONAL STATUS ASSESSMENT

Full nutritional status assessment is reserved for seriously ill patients and those at very high nutritional risk when the cause of malnutrition is still uncertain after the initial clinical evaluation and dietary assessment. It involves multiple dimensions, including documentation of dietary intake, anthropometric measurements, biochemical measurements of blood and urine, clinical examination, health history elicitation, and functional status evaluation. Therapeutic dietary prescriptions and menu plans for most diseases are available from most hospitals and from the Academy of Nutrition and Dietetics. [For further discussion of nutritional assessment, see Chap. 334.](#)

GLOBAL CONSIDERATIONS

The DRIs (e.g., the EAR, the UL, and energy needs) are estimates of physiologic requirements based on experimental evidence. Assuming that appropriate adjustments are made for age, sex, body size, and physical activity level, these estimates should be applicable to individuals in most parts of the world. However, other values are not transportable. The AIs are based on customary and adequate intakes in U.S. and Canadian populations, which appear to be compatible with good health, rather than on a large body of direct experimental evidence. Similarly, the AMDRs represent expert opinion regarding the approximate intakes of energy-providing nutrients that are healthful in these North American populations, and the CDRR may also vary in other populations. Thus, these measures should be used with caution in other settings. Nutrient-based standards like the DRIs have also been developed by the World Health Organization/Food and Agricultural Organization of the United Nations and are available on the Web (<https://www.who.int/activities/establishing-global-nutrient-requirements>). The European Food Safety Authority (EFSA) Panel on Dietetic Products, Nutrition, and Allergies periodically publishes its recommendations in the online *EFSA Journal* (<https://efsa.onlinelibrary.wiley.com/journal/18314732>). Other countries have promulgated similar recommendations. The different standards have many similarities in their basic concepts, definitions, and nutrient recommendation levels, but there are some differences from the DRIs as a result of the functional criteria chosen, environmental differences, the timeliness of the evidence reviewed, and expert judgment. There is a growing trend toward global harmonization of these recommendations.

FURTHER READING

- B PM et al: Scanning for new evidence to prioritize updates to the Dietary Reference Intakes: Case studies for thiamin and phosphorus. *Am J Clin Nur* 104:1366, 2016.
- F H et al: Personalized nutrition: The role of new dietary assessment methods. *Proc Nutr Soc* 75:96, 2016.
- G RS: *Principles of Nutritional Assessment*, 2nd ed. Oxford, Oxford University Press, 2005.
- L JL, D JT: Establishing nutrient intake values, in *Present Knowledge in Nutrition*, Vol 2. BP Marriott, DF Birt, VA Stallings, AA Yates, eds. London, Academic Press, 2020, pp 267–289.

M BP et al: Present knowledge, in *Nutrition Vol 1: Basic Nutrition and Metabolism*, Vol 2: *Clinical and Applied Topics in Nutrition*. London, Academic Press, 2020.

N A S , E , M : Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease. Washington DC, National Academies Press, 2017.

N A S , E , M : Global Harmonization of Methodological Approaches to Nutrient Intake Recommendations: Proceedings of a Workshop. Washington DC, National Academies Press, 2018.

N A S , E , M : Dietary Reference Intakes for Sodium and Potassium. Washington DC, National Academies Press, 2019.

N A S , E , M : Advancing Nutrition and Food Science: 80th Anniversary of the Food and Nutrition Board: Proceedings of a Symposium. Washington DC, National Academies Press, 2020.

N A S , E , M : Harmonizing the Process for Establishing Nutrient Reference Values: A Tool Kit. Washington DC, National Academies Press, 2020.

R	S	I	U	R	L
D	R	I	U	R	L
N	,	S	C		S
E		D	R	I	, F N -

B : Dietary Reference Intakes: Applications in Dietary Assessment. Washington, DC, National Academies Press, 2008.

S PJ, K JC: More nutrition precision, better decisions for the health of our nation *J Nutr* 150:3058, 2020.

Y A et al: Why the derivation of nutrient reference values should be harmonized and how it can be accomplished. *Adv Nutr* 11:1112, 2020.

Y EA et al: Options for basing Dietary Reference Intakes (DRIs) on chronic disease endpoints report from a joint US-/Canadian-sponsored working group. *Am J Clin Nutr* 105:249S, 2017.

333

Vitamin and Trace Mineral Deficiency and Excess

Paolo M. Suter



Vitamins are required constituents of the human diet because they are synthesized inadequately or not at all in the human body. Only small amounts of these substances are needed to carry out essential biochemical reactions (e.g., by acting as coenzymes or prosthetic groups). Overt vitamin or trace mineral deficiencies are rare in Western countries because of a plentiful, varied, and inexpensive food supply; food fortification; and use of supplements. However, multiple nutrient deficiencies may appear together in persons who are chronically ill, alcoholic, or living in poverty. After bariatric surgery, patients are at high risk for multiple nutrient deficiencies. Moreover, subclinical vitamin and trace mineral deficiencies (often designated as “hidden hunger”), as diagnosed by laboratory testing, are quite common in the normal population, especially in the geriatric age group and socioeconomically deprived individuals due to the lack of nutrient-dense foods. Conversely, because of the widespread use of nutrient supplements and food fortification, nutrient toxicities are gaining pathophysiologic and clinical importance.

Victims of famine, emergency-affected and displaced populations, refugees, and camp populations are at increased risk for protein-energy malnutrition and classic micronutrient deficiencies (vitamin A, iron,

Body stores of vitamins and minerals vary tremendously. For example, stores of vitamins B₁₂ and A are large, and an adult may not become deficient until ≥1 year after beginning to eat a deficient diet. However, folate and thiamine may become depleted within weeks among those eating a deficient diet. Therapeutic modalities can deplete essential nutrients from the body; for example, hemodialysis or diuretics remove water-soluble vitamins, which must be replaced by supplementation.

Vitamins and trace minerals play several roles in diseases: (1) deficiencies of vitamins and minerals may be caused by disease states such as malabsorption; (2) either deficiency or excess of vitamins and minerals can cause disease in and of itself (e.g., vitamin A intoxication and liver disease); and (3) vitamins and minerals in high doses may be used as drugs (e.g., niacin for hypercholesterolemia). Since they are covered elsewhere, the hematologic-related vitamins and minerals (Chaps. 97 and 99) either are not considered or are considered only briefly in this chapter, as are the bone-related vitamins and minerals (vitamin D, calcium, phosphorus, magnesium; Chap. 409).

VITAMINS

See also Table 333-1 and Fig. 333-1.

■ THIAMINE VITAMIN B₁

Thiamine was the first B vitamin to be identified and therefore is referred to as vitamin B₁. Thiamine functions in the decarboxylation of α-ketoacids (e.g., pyruvate α-ketoglutarate) and branched-chain amino acids and thus is essential for energy generation. In addition, thiamine pyrophosphate acts as a coenzyme for a transketolase reaction that mediates the conversion of hexose and pentose phosphates. It has been postulated that thiamine plays a role in peripheral nerve conduction, although the exact chemical reactions underlying this function are not known.

Food Sources The median intake of thiamine in the United States from food alone is ~2 mg/d. Primary food sources for thiamine include yeast, organ meat, pork, legumes, beef, whole grains, and nuts. Milled rice and grains contain little thiamine. Thiamine deficiency is therefore more common in cultures that rely heavily on a milled polished rice-based diet. Certain foods contain antithiamine factors such as

heat-labile thiaminases (raw fish, shellfish), which destroy the vitamin, or heat-stable polyhydroxyphenols (tannins; in coffee, tea, Brussels sprouts, or betel nuts), which inactivate the vitamin. Thus, drinking large amounts of tea or coffee could theoretically lower thiamine body stores.

Deficiency Most dietary deficiency of thiamine worldwide is the result of poor dietary intake due to the lack of food or disproportionate reliance on highly processed staple crops. Food processing removes thiamine, and high-heat or long-duration cooking destroys it. In Western countries, the primary causes of thiamine deficiency are alcoholism and chronic illnesses such as cancer. Alcohol interferes directly with the absorption of thiamine and with the synthesis of thiamine pyrophosphate, and it increases urinary excretion. Thiamine should always be replenished when a patient with alcoholism is being refed, as carbohydrate repletion without adequate thiamine can precipitate acute thiamine deficiency with lactic acidosis. Other at-risk populations are women with prolonged hyperemesis gravidarum and anorexia, patients with overall poor nutritional status who are receiving parenteral glucose, patients who have had bariatric/metabolic surgery (*bariatric Wernicke*), and patients receiving chronic diuretic therapy (e.g., in hypertension or systolic heart failure) due to increased urinary thiamine losses. Different drugs (e.g., metformin, verapamil) could inhibit intestinal thiamine transporters (ThTR-2), thereby increasing the risk of deficiency for this vitamin. Maternal thiamine deficiency can lead to infantile beriberi in breast-fed children. Thiamine deficiency could be an underlying factor in motor vehicle accidents and could be overlooked in the setting of head injury.

Thiamine deficiency in its early stage induces anorexia and non-specific symptoms (e.g., irritability, decrease in short-term memory). Prolonged thiamine deficiency causes *beriberi*, which is classically categorized as wet or dry, although there is considerable overlap between the two categories. In either form of beriberi, patients may complain of pain and paresthesia. *Wet beriberi* presents primarily with cardiovascular symptoms that are due to impaired myocardial energy metabolism and dysautonomia; it can occur after 3 months of a thiamine-deficient diet. Patients present with an enlarged heart, tachycardia, high-output congestive heart failure, peripheral edema, and peripheral neuritis. Patients with *dry beriberi* present with a symmetric

TABLE 333-1 Principal Clinical Findings of Vitamin Malnutrition

NUTRIENT	CLINICAL FINDING	DIETARY LEVEL PER DAY ASSOCIATED WITH OVERT DEFICIENCY IN ADULTS	CONTRIBUTING FACTORS TO DEFICIENCY
Thiamine	Beriberi: neuropathy, muscle weakness and wasting, cardiomegaly, edema, ophthalmoplegia, confabulation	<0.3 mg/1000 kcal	Alcoholism, chronic diuretic use, bariatric surgery, hyperemesis, thiaminases in food
Riboflavin	Magenta tongue, angular stomatitis, seborrhea, cheilosis, ocular symptoms, corneal vascularization	<0.4 mg	Alcoholism, individuals with poor diets and low intake of milk products
Niacin	Pellagra: pigmented rash of sun-exposed areas, bright red tongue, diarrhea, apathy, memory loss, disorientation	<9.0 niacin equivalents	Alcoholism, vitamin B ₃ deficiency, riboflavin deficiency, tryptophan deficiency
Vitamin B ₆	Seborrhea, glossitis, convulsions, neuropathy, depression, confusion, microcytic anemia	<0.2 mg	Alcoholism, isoniazid
Folate	Megaloblastic anemia, atrophic glossitis, depression, ↑ homocysteine	<100 µg/d	Alcoholism, sulfasalazine, pyrimethamine, triamterene
Vitamin B ₁₂	Megaloblastic anemia, loss of vibratory and position sense, abnormal gait, dementia, impotence, loss of bladder and bowel control, ↑ homocysteine, ↑ methylmalonic acid	<1.0 µg/d	Gastric atrophy (pernicious anemia), terminal ileal disease, strict vegetarianism, acid-reducing drugs (e.g., H ₂ blockers), metformin
Vitamin C	Scurvy: petechiae, ecchymosis, coiled hairs, inflamed and bleeding gums, joint effusion, poor wound healing, fatigue	<10 mg/d	Smoking, alcoholism
Vitamin A	Xerophthalmia, night blindness, Bitot's spots, follicular hyperkeratosis, impaired embryonic development, immune dysfunction	<300 µg/d	Fat malabsorption, infection, measles, alcoholism, protein-energy malnutrition
Vitamin D	Rickets: skeletal deformation, rachitic rosary, bowed legs; osteomalacia	<2.0 µg/d	Aging, lack of sunlight exposure, fat malabsorption, deeply pigmented skin
Vitamin E	Peripheral neuropathy, spinocerebellar ataxia, skeletal muscle atrophy, retinopathy	Not described unless underlying contributing factor is present	Occurs only with fat malabsorption or genetic abnormalities of vitamin E metabolism/transport
Vitamin K	Elevated prothrombin time, bleeding	<10 µg/d	Fat malabsorption, liver disease, antibiotic use

peripheral neuropathy of the motor and sensory systems, with diminished reflexes. The neuropathy affects the legs most markedly, and patients have difficulty rising from a squatting position.

Alcoholic patients with chronic thiamine deficiency also may have central nervous system (CNS) manifestations known as *Wernicke's encephalopathy*, which consists of horizontal nystagmus, ophthalmoplegia (due to weakness of one or more extraocular muscles), cerebellar ataxia, and mental impairment (Chap. 453). When there

is an additional loss of memory and a confabulatory psychosis, the syndrome is known as *Wernicke-Korsakoff syndrome*. Despite the typical clinical picture and history, Wernicke-Korsakoff syndrome is underdiagnosed.

The laboratory diagnosis of thiamine deficiency usually is made by a functional enzymatic assay of transketolase activity measured before and after the addition of thiamine pyrophosphate. A >25% stimulation in response to the addition of thiamine pyrophosphate (i.e., an activity

Vitamin	Active derivative or cofactor form	Principal function
Thiamine (B_1) 	Thiamine pyrophosphate	Coenzyme for cleavage of carbon-carbon bonds; amino acid and carbohydrate metabolism
Riboflavin (B_2) 	Flavin mononucleotide (FMN) and flavin adenine dinucleotide (FAD)	Cofactor for oxidation, reduction reactions, and covalently attached prosthetic groups for some enzymes
Niacin 	Nicotinamide adenine dinucleotide phosphate (NADP) and nicotinamide adenine dinucleotide (NAD)	Coenzymes for oxidation and reduction reactions
Vitamin B_6 	Pyridoxal phosphate	Cofactor for enzymes of amino acid metabolism
Folate 	Polyglutamate forms of (5, 6, 7, 8) tetrahydrofolate with carbon unit attachments	Coenzyme for one carbon transfer in nucleic acid and amino acid metabolism
Vitamin B_{12} 	Methylcobalamin Adenosylcobalamin	Coenzyme for methionine synthase and L-methylmalonyl-CoA mutase

FIGURE 333-1 Structures and principal functions of vitamins associated with human disorders.

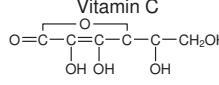
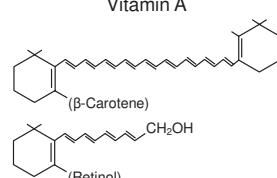
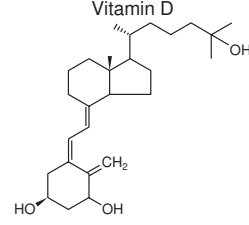
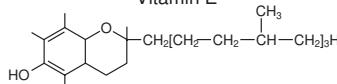
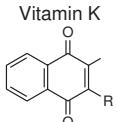
Vitamin	Active derivative or cofactor form	Principal function
Vitamin C 	Ascorbic acid and dehydroascorbic acid	Participation as a redox ion in many biologic oxidation and hydrogen transfer reactions
Vitamin A 	Retinol, retinaldehyde, and retinoic acid	Formation of rhodopsin (vision) and glycoproteins (epithelial cell function); also regulates gene transcription
Vitamin D 	1,25-Dihydroxyvitamin D	Maintenance of blood calcium and phosphorus levels; antiproliferative hormone
Vitamin E 	Tocopherols and tocotrienols	Antioxidants
Vitamin K 	Vitamin K hydroquinone	Cofactor for posttranslational carboxylation of many proteins including essential clotting factors

FIGURE 333-1 (Continued)

coefficient of 1.25) is interpreted as abnormal. Thiamine or the phosphorylated esters of thiamine in serum or blood also can be measured by high-performance liquid chromatography to detect deficiency.

TREATMENT

Thiamine Deficiency

In acute thiamine deficiency with either cardiovascular or neurologic signs, 200 mg of thiamine three times daily should be given intravenously until there is no further improvement in acute symptoms; oral thiamine (10 mg/d) should subsequently be given until recovery is complete. Cardiovascular and ophthalmoplegic improvement occurs within 24 h. Other manifestations gradually clear, although psychosis in Wernicke-Korsakoff syndrome may be permanent or may persist for several months. Other nutrient deficiencies should be corrected concomitantly. In view of the widespread, often unrecognized (subclinical) deficiency, a more generous supplementation of this vitamin in the emergency care setting is warranted.

Toxicity Although hypersensitivity/anaphylaxis has been reported after high intravenous doses of thiamine, no adverse effects have been recorded from either food or supplements at high doses.

RIBOFLAVIN VITAMIN B₂

Riboflavin is important for the metabolism of fat, carbohydrate, and protein, acting as a respiratory coenzyme and an electron donor.

Enzymes that contain flavin adenine dinucleotide (FAD) or flavin mononucleotide (FMN) as prosthetic groups are known as *flavoenzymes* (e.g., succinic acid dehydrogenase, monoamine oxidase, glutathione reductase). FAD is a cofactor for methyltetrahydrofolate reductase and therefore modulates homocysteine metabolism. The vitamin also plays a role in drug and steroid metabolism, including detoxification reactions.

Although much is known about the chemical and enzymatic reactions of riboflavin, the clinical manifestations of riboflavin deficiency are nonspecific and are similar to those of other deficiencies of B vitamins. Riboflavin deficiency is manifested principally by lesions of the mucocutaneous surfaces of the mouth and skin. In addition, corneal vascularization, anemia, and personality changes have been described with riboflavin deficiency.

Deficiency and Excess Riboflavin deficiency almost always is due to dietary deficiency. Milk, other dairy products, and enriched breads and cereals are the most important dietary sources of riboflavin in the United States, although lean meat, fish, eggs, broccoli, and legumes are also good sources. Riboflavin is extremely sensitive to light, and milk should be stored in containers that protect against photodegradation. Laboratory diagnosis of riboflavin deficiency can be made by determination of red blood cell or urinary riboflavin concentrations or by measurement of erythrocyte glutathione reductase activity, with and without added FAD. Because of the limited capacity of the gastrointestinal tract to absorb riboflavin (~27 mg after one oral dose) as well as the instantaneous urinary excretion, riboflavin toxicity has not been described.

■ NIACIN VITAMIN B₃

The term *niacin* refers to nicotinic acid and nicotinamide and their biologically active derivatives. Nicotinic acid and nicotinamide serve as precursors of two coenzymes, nicotinamide adenine dinucleotide (NAD) and NAD phosphate (NADP), which are important in numerous oxidation and reduction reactions in the body. In addition, NAD and NADP are active in adenine diphosphate–ribose transfer reactions involved in DNA repair and calcium mobilization.

Metabolism and Requirements Nicotinic acid and nicotinamide are absorbed well from the stomach and small intestine. The bioavailability of niacin from beans, milk, meat, and eggs is high; bioavailability from cereal grains is lower. Since flour is enriched with “free” niacin (i.e., the non-coenzyme form), bioavailability is excellent. Median intakes of niacin in the United States considerably exceed the recommended dietary allowance (RDA).

The amino acid tryptophan can be converted to niacin with an efficiency of 60:1 by weight. Thus, the RDA for niacin is expressed in niacin equivalents. A lower-level conversion of tryptophan to niacin occurs in vitamin B₆ and/or riboflavin deficiencies and in the presence of isoniazid. The urinary excretion products of niacin include 2-pyridone and 2-methyl nicotinamide, measurements of which are used in the diagnosis of niacin deficiency.

Deficiency Niacin deficiency causes *pellagra*, which is found mostly among people eating corn-based diets in parts of China, Africa, and India. Pellagra in North America is found mainly among alcoholics; among patients with congenital defects of intestinal and kidney absorption of tryptophan (Hartnup disease; *Chap. 420*); and among patients with carcinoid syndrome (*Chap. 84*), in which there is increased conversion of tryptophan to serotonin. The antituberculosis drug isoniazid is a structural analogue of niacin and can precipitate pellagra. In the setting of famine or population displacement, pellagra results from the absolute lack of niacin but also from the deficiency of micronutrients required for the conversion of tryptophan to niacin (e.g., iron, riboflavin, and pyridoxine). The early symptoms of pellagra include loss of appetite, generalized weakness and irritability, abdominal pain, and vomiting. Bright red glossitis then ensues and is followed by a characteristic skin rash that is pigmented and scaling, particularly in skin areas exposed to sunlight. This rash is known as *Casal's necklace* because it forms a ring around the neck; it is seen in advanced cases. Vaginitis and esophagitis also may occur. Diarrhea (due in part to proctitis and in part to malabsorption), depression, seizures, and dementia are also part of the pellagra syndrome. The primary manifestations of this syndrome are sometimes referred to as “the four Ds”: dermatitis, diarrhea, and dementia leading to death. Aging is characterized by a decline in cellular NAD⁺, and it seems plausible that maintaining and/or reestablishing cellular NAD⁺, might favorably modulate the risk of chronic diseases of aging (e.g., metabolic disorders).

TREATMENT

Pellagra

Treatment of pellagra consists of oral supplementation with 100–200 mg of nicotinamide or nicotinic acid three times daily for 5 days. High doses of nicotinic acid (2 g/d in a time-release form) are used for the treatment of elevated cholesterol and triglyceride levels and/or low high-density lipoprotein cholesterol levels, but without proven evidence to prevent cardiovascular disease. Nevertheless, nicotinic acid may be useful in patients with statin intolerance or severe hypertriglyceridemia (*Chap. 407*).

Toxicity Prostaglandin-mediated flushing due to binding of the vitamin to a G protein-coupled receptor has been observed at daily nicotinic acid doses as low as 30 mg taken as a supplement or as therapy for dyslipidemia. There is no evidence of toxicity from niacin that is derived from food sources. Flushing always starts in the face and may be accompanied by skin dryness, itching, paresthesia, and

headache. Flushing is subject to tachyphylaxis and often improves with time; premedication with aspirin may alleviate these symptoms. Nausea, vomiting, and abdominal pain also occur at similar doses of niacin. Hepatic toxicity is the most serious toxic reaction caused by sustained-release niacin and may present as jaundice with elevated aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels. A few cases of fulminant hepatitis requiring liver transplantation have been reported at doses of 3–9 g/d. Other toxic reactions include glucose intolerance, hyperuricemia, macular edema, and macular cysts. The combination of nicotinic acid preparations for dyslipidemia plus 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase inhibitors may increase the risk of rhabdomyolysis. The upper limit for daily (nontherapeutic) niacin intake has been set at 35 mg.

■ PYRIDOXINE VITAMIN B₆

Vitamin B₆ refers to a family of compounds that includes pyridoxine, pyridoxal, pyridoxamine, and their 5'-phosphate derivatives. 5'-Pyridoxal phosphate (PLP) is a cofactor for >100 enzymes involved in amino acid metabolism. Vitamin B₆ also is involved in heme and neurotransmitter synthesis and in the metabolism of glycogen, lipids, steroids, sphingoid bases, and several vitamins, including the conversion of tryptophan to niacin.

Dietary Sources Plants contain vitamin B₆ in the form of pyridoxine, whereas animal tissues contain PLP and pyridoxamine phosphate. The vitamin B₆ contained in plants is less bioavailable than that in animal tissues. Rich food sources of vitamin B₆ include legumes, nuts, wheat bran, and meat, although it is present in all food groups.

Deficiency Symptoms of vitamin B₆ deficiency include epithelial changes, as seen frequently with other B vitamin deficiencies. In addition, severe vitamin B₆ deficiency can lead to peripheral neuropathy, abnormal electroencephalograms, and personality changes that include depression and confusion. In infants, diarrhea, seizures, and anemia have been reported. Microcytic hypochromic anemia is due to diminished hemoglobin synthesis, since the first enzyme involved in heme biosynthesis (aminolevulinate synthase) requires PLP as a cofactor (*Chap. 97*). In some case reports, platelet dysfunction has been reported. Since vitamin B₆ is necessary for the conversion of homocysteine to cystathione, it is possible that chronic low-grade vitamin B₆ deficiency may result in hyperhomocysteinemia, which has been associated with vascular dysfunction and an increased risk of cardiovascular disease; however, so far, there is only limited randomized controlled trial evidence (*Chap. 420*). Independent of homocysteine, low levels of circulating vitamin B₆ have been associated with inflammation and elevated levels of C-reactive protein.

Certain medications, such as isoniazid, -dopa, penicillamine, and cycloserine, interact with PLP due to a reaction with carbonyl groups. Pyridoxine should be given concurrently with isoniazid to avoid neuropathy. The increased ratio of AST to ALT seen in alcoholic liver disease reflects the relative vitamin B₆ dependence of ALT. Vitamin B₆ dependency syndromes that require pharmacologic doses of vitamin B₆ are rare; they include cystathione β -synthase deficiency, pyridoxine-responsive (primarily sideroblastic) anemias, and gyrate atrophy with chorioretinal degeneration due to decreased activity of the mitochondrial enzyme ornithine aminotransferase. In these situations, 100–200 mg/d of oral vitamin B₆ is required for treatment.

Severe nausea and vomiting in pregnancy might respond to pyridoxine combined with doxylamine. High doses of vitamin B₆ have been used to treat carpal tunnel syndrome, premenstrual syndrome, schizophrenia, autism, and diabetic neuropathy but have not been found to be effective.

The laboratory diagnosis of vitamin B₆ deficiency is generally based on low plasma PLP values (<20 nmol/L). Vitamin B₆ deficiency is treated with 50 mg/d; higher doses of 100–200 mg/d are given if the deficiency is related to medication use. Vitamin B₆ should not be given with -dopa, since the vitamin interferes with the action of this drug.

Toxicity The safe upper limit for vitamin B₆ has been set at 100 mg/d, although no adverse effects have been associated with high intakes of

2528 vitamin B₆ from food sources only. When toxicity occurs, it causes severe sensory neuropathy, leaving patients unable to walk; however, in most cases, this is reversible upon cessation of the high intake. Medication safety monitoring suggests a rather high prevalence of vitamin B₆-induced neuropathy. Accordingly, long-term high-dose vitamin B₆ supplementation should be discouraged. Some cases of photosensitivity and dermatitis have been reported.

■ FOLATE VITAMIN B₁₂

See Chap. 99.

■ VITAMIN C

Both ascorbic acid (only the -isomer) and its oxidized product dehydroascorbic acid are biologically active. Actions of vitamin C include antioxidant activity, promotion of nonheme iron absorption, carnitine biosynthesis, conversion of dopamine to norepinephrine, tyrosine catabolism, histone and DNA demethylation, and synthesis of many peptide hormones. Vitamin C is also important for connective tissue metabolism and cross-linking (proline hydroxylation), and it is a component of many drug-metabolizing enzyme systems, particularly the mixed-function oxidase systems.

Absorption and Dietary Sources Vitamin C is almost completely absorbed if <100 mg is administered in a single dose; however, only $\leq 50\%$ is absorbed at doses >1 g. Enhanced degradation and fecal and urinary excretion of vitamin C occur at higher intake levels.

Good dietary sources of vitamin C include citrus fruits, green vegetables (especially broccoli), tomatoes, and potatoes. Consumption of five servings of fruits and vegetables a day provides vitamin C in excess of the RDA of 90 mg/d for men and 75 mg/d for women. In addition, ~40% of the U.S. population consumes vitamin C as a dietary supplement in which "natural forms" of the vitamin are no more bioavailable than synthetic forms. Smoking (including "passive" smoking), hemodialysis, pregnancy, lactation, and stress (e.g., infection, trauma) appear to increase vitamin C requirements.

Deficiency Vitamin C deficiency causes scurvy. In the United States, this condition is seen primarily among the poor and the elderly, in alcoholics who consume <10 mg/d of vitamin C, and in young adults who eat severely unbalanced diets. In addition to generalized fatigue, symptoms of scurvy primarily reflect impaired formation of mature connective tissue and include bleeding into the skin (petechiae, ecchymoses, perifollicular hemorrhages); inflamed and bleeding gums; and manifestations of bleeding into joints, the peritoneal cavity, the pericardium, and the adrenal glands. In children, vitamin C deficiency may cause impaired bone growth. Laboratory diagnosis of vitamin C deficiency is based on low plasma or leukocyte levels.

Administration of vitamin C (200 mg/d) improves the symptoms of scurvy within several days. High-dose vitamin C supplementation (e.g., 0.2 g up to several grams per day) may slightly decrease the symptoms and duration of upper respiratory tract infections. Vitamin C supplementation has also been reported to be useful in Chédiak-Higashi syndrome (Chap. 64) and osteogenesis imperfecta (Chap. 413). Diets high in vitamin C have been claimed to lower the incidence of certain cancers, particularly esophageal and gastric cancers. If proven, this effect may be because vitamin C can prevent the conversion of nitrates and secondary amines to carcinogenic nitrosamines. Emerging evidence suggests a therapeutic effect of intravenous parenteral (*not oral*) pharmacologic doses of up to 1 g/kg body weight of ascorbic acid in the treatment of cancers (e.g., metastatic pancreatic, ovarian, glioblastoma, and non-small-cell lung cancers). The mechanism of pharmacologic ascorbate in cancer treatment (as a stand-alone agent or with other therapeutic agents) appears to be pro-oxidative, either synergistic (e.g., gemcitabine, PD-1 inhibitors, radiation) or additive with other agents.

Toxicity Taking >2 g of vitamin C in a single dose may result in abdominal pain, diarrhea, and nausea. Since vitamin C may be metabolized to oxalate, it is feared that chronic high-dose vitamin C supplementation could result in an increased prevalence of kidney stones. However, except in patients with preexisting renal disease, this association has not been borne out in several trials. Nevertheless, it is

reasonable to advise patients with a history of kidney stones (especially oxalate renal stones) and renal insufficiency not to take large doses of vitamin C. There is also an unproven but possible risk that chronic high doses of vitamin C could promote iron overload and iron toxicity (e.g., in patients with hemochromatosis or thalassemia major). High doses of vitamin C can induce hemolysis in patients with glucose-6-phosphate dehydrogenase deficiency, and doses >1 g/d can cause false-negative guaiac reactions and interfere with tests for urinary glucose. High doses may interfere with the activity of certain drugs and diagnostic tests (e.g., false-negative results of guaiac-based fecal occult blood tests).

■ BIOTIN

Biotin (also known as vitamin B₇ or vitamin H) is a water-soluble vitamin that plays a role in gene expression, gluconeogenesis, and fatty acid synthesis and serves as a carbon dioxide (CO₂) carrier on the surface of both cytosolic and mitochondrial carboxylase enzymes. The vitamin also functions in the catabolism of specific amino acids (e.g., leucine) and in gene regulation by histone biotinylation. Excellent food sources of biotin include organ meat such as liver or kidney, soy and other beans, yeast, and egg yolks; however, egg white contains the protein avidin, which strongly binds the vitamin and reduces its bioavailability.

Biotin deficiency due to low dietary intake is rare; rather, deficiency is due to inborn errors of metabolism. Biotin deficiency has been induced by experimental feeding of egg white diets and by biotin-free parenteral nutrition in patients with short bowels. In adults, biotin deficiency results in mental changes (depression, hallucinations), paresthesia, anorexia, and nausea. A scaling, seborrheic, and erythematous rash may occur around the eyes, nose, and mouth as well as on the extremities. In infants, biotin deficiency presents as hypotonia, lethargy, and apathy. In addition, infants may develop alopecia and a characteristic rash that includes the ears. At present, evidence does not support a therapeutic role of high-dose biotin in multiple sclerosis. The laboratory diagnosis of biotin deficiency can be established on the basis of a decreased concentration of urinary biotin (or its major metabolites), increased urinary excretion of 3-hydroxyisovaleric acid after a leucine challenge, or decreased activity of biotin-dependent enzymes in lymphocytes (e.g., propionyl-CoA carboxylase). Treatment requires pharmacologic doses of biotin, that is, up to 10 mg/d. No toxicity is known. High-dose biotin supplements could interfere with different immunoassay platforms based on streptavidin-biotin technology (e.g., biotinylated immunoassays), resulting in false-positive (e.g., free T₄ or T₃) or false-negative tests (e.g., thyroid-stimulating hormone, troponin, β -human chorionic gonadotropin pregnancy test).

■ PANTOTHENIC ACID VITAMIN B₅

Pantothenic acid is a component of coenzyme A and phosphopantethine, which are involved in fatty acid metabolism and the synthesis of cholesterol, steroid hormones, and all compounds formed from isoprenoid units. In addition, pantothenic acid is involved in the acetylation of proteins. The vitamin is excreted in the urine, and the laboratory diagnosis of deficiency is based on low urinary vitamin levels.

The vitamin is ubiquitous in the food supply. Liver, yeast, egg yolks, whole grains, and vegetables are particularly good sources. Human pantothenic acid deficiency has been demonstrated only by experimental feeding of diets low in pantothenic acid or by administration of a specific pantothenic acid antagonist. The symptoms of pantothenic acid deficiency are nonspecific and include gastrointestinal disturbance, depression, muscle cramps, paresthesia, ataxia, and hypoglycemia. Pantothenic acid deficiency is believed to have caused the "burning feet syndrome" seen in prisoners of war during World War II. No toxicity of this vitamin has been reported.

■ CHOLINE

Choline is a precursor for acetylcholine, phospholipids, and betaine. Choline is necessary for the structural integrity of cell membranes, cholinergic neurotransmission, lipid and cholesterol metabolism, methyl-group metabolism, and transmembrane signaling. Recently, a recommended adequate intake was set at 550 mg/d for men and

425 mg/d for women, although certain genetic polymorphisms can increase an individual's requirement. Choline is thought to be a "conditionally essential" nutrient in that its de novo synthesis occurs in the liver and results in lesser-than-used amounts only under certain stress conditions (e.g., alcoholic liver disease). The dietary requirement for choline depends on the status of other nutrients involved in methyl-group metabolism (folate, vitamin B₁₂, vitamin B₆, and methionine) and thus varies widely. Choline is widely distributed in food (e.g., egg yolks, wheat germ, organ meat, milk) in the form of lecithin (phosphatidylcholine). Choline deficiency has occurred only in experimental conditions or in patients receiving parenteral nutrition devoid of choline and rarely in specific inborn errors of choline metabolism. Deficiency results in fatty liver, elevated aminotransferase levels, and skeletal muscle damage with high creatine phosphokinase values. The diagnosis of choline deficiency is currently based on low plasma levels, although nonspecific conditions (e.g., heavy exercise) may also suppress plasma levels.

Toxicity from choline results in hypotension, cholinergic sweating, diarrhea, salivation, and a fishy body odor. The upper limit for choline intake has been set at 3.5 g/d. Because of its ability to lower cholesterol and homocysteine levels, choline treatment has been suggested for patients with dementia and patients at high risk of cardiovascular disease. However, the benefits of such treatment have not been firmly documented; recently, signals for an increased cardiovascular risk have been reported. Choline- and betaine-restricted diets are of therapeutic value in trimethylaminuria ("fish odor syndrome") or in decreasing the production of the gut microbiome-derived trimethylamine N-oxide (TMAO) as a potential cardiovascular risk modulator.

■ FLAVONOIDS

Flavonoids constitute a large family of polyphenols that contribute to the aroma, taste, and color of fruits and vegetables. Major groups of dietary flavonoids include anthocyanidins in berries; catechins in green tea and chocolate; flavonols (e.g., quercetin) in broccoli, kale, leeks, onions, and the skins of grapes and apples; and isoflavones (e.g., genistein) in legumes. Isoflavones have a low bioavailability and are partially metabolized by the intestinal flora. The dietary intake of flavonoids is estimated at 10–100 mg/d; this figure is almost certainly an underestimate attributable to a lack of information on their concentrations in many foods. Several flavonoids have antioxidant activity and affect cell signaling. From observational epidemiologic studies and limited clinical (human and animal) studies, flavonoids have been postulated to play a role in the prevention of several chronic diseases, including neurodegenerative disease, diabetes, and osteoporosis. The ultimate importance and usefulness of these compounds against human disease have not been consistently demonstrated. Nevertheless, a dietary pattern with high intake of fruits, vegetables, and legumes should be encouraged to assure a higher intake of these and other nonnutritive bioactives.

■ VITAMIN A

Vitamin A, in the strictest sense, refers to retinol and retinyl esters. However, the oxidized metabolites retinaldehyde and retinoic acid are also biologically active compounds. The term *retinoids* includes all molecules (including synthetic molecules) that are chemically related to retinol. Retinaldehyde (11-cis) is the form of vitamin A that is required for normal vision, whereas retinoic acid is necessary for normal morphogenesis, growth, and cell differentiation. Retinoic acid does not function directly in vision and, in contrast to retinol, is not involved in reproduction. Vitamin A also plays a role in iron utilization, humoral immunity, T cell-mediated immunity, natural killer cell activity, and phagocytosis.

Vitamin A is found in the human food supply in two forms: preformed as retinyl esters and provitamin A carotenoids. There are >700 carotenoids in nature, ~50 of which can be metabolized to vitamin A. β-Carotene is the most prevalent carotenoid with provitamin A activity in the food supply. In humans, significant fractions of carotenoids are absorbed intact and are stored in liver and fat. It is estimated that in healthy humans ≥12 µg (range, 4–27 µg) of dietary all-trans β-carotene

is equivalent to 1 µg of retinol activity, whereas the figure is ≥24 µg for other dietary provitamin A carotenoids (e.g., β-cryptoxanthin, α-carotene). The vitamin A equivalency for a β-carotene supplement in an oily solution is 2:1.

Metabolism The liver contains ~90% of the vitamin A reserves in healthy individuals and secretes vitamin A in the form of retinol, which is bound in the circulation to retinol-binding protein. Once binding has occurred, the retinol-binding protein complex interacts with a second protein, transthyretin. This trimolecular complex functions to prevent vitamin A from being filtered by the kidney glomerulus, thus protecting the body against the toxicity of retinol and allowing retinol to be taken up by specific cell-surface receptors that recognize retinol-binding protein. A certain amount of vitamin A enters peripheral cells even if it is not bound to retinol-binding protein. After retinol is internalized by the cell, it becomes bound to a series of cellular retinol-binding proteins, which function as sequestering and transporting agents as well as co-ligands for enzymatic reactions. Certain cells also contain retinoic acid–binding proteins, which have sequestering functions but also shuttle retinoic acid to the nucleus and enable its metabolism.

Vitamin A metabolites (retinoids) such as retinoic acid are potent regulators of gene transcription through nuclear receptor signaling, thus playing a key role in many cellular and metabolic pathways. Two families of receptors (retinoic acid receptors [RARs] and retinoid X receptors [RXRs]) are active in retinoid-mediated gene transcription. Retinoid receptors regulate transcription by binding as dimeric complexes to specific DNA sites—the retinoic acid response elements—in target genes (Chap. 377). The receptors can either stimulate or repress gene expression in response to their ligands. RARs bind all-trans retinoic acid and 9-cis-retinoic acid, whereas RXRs bind only 9-cis-retinoic acid.

The retinoid receptors play an important role in controlling cell proliferation and differentiation. RXRs dimerize with other nuclear receptors to function as coregulators of genes responsive to retinoids, but also to thyroid hormone and calcitriol. RXR agonists induce insulin sensitivity experimentally, perhaps because RXRs are cofactors for the peroxisome proliferator-activated receptors, which also mediate fatty acid and carbohydrate metabolism and are targets for different drugs including thiazolidinedione drugs (e.g., rosiglitazone and pioglitazone) (Chap. 404).

Dietary Sources The retinol activity equivalent (RAE) is used to express the vitamin A value of food: 1 RAE is defined as 1 µg of retinol (0.003491 mmol), 12 µg of β-carotene, and 24 µg of other provitamin A carotenoids. In older literature, vitamin A often was expressed in international units (IUs), with 1 µg of retinol equal to 3.33 IU of retinol and 20 IU of β-carotene. Although these IUs are no longer in scientific use, they can still be found in reports of the food industry and in public health interventions in low-income countries.

Liver, fish, and eggs are excellent food sources for preformed vitamin A; vegetable sources of provitamin A carotenoids include dark green and deeply colored fruits and vegetables. Moderate cooking of vegetables enhances carotenoid release for uptake in the gut. Carotenoid absorption is also aided by some fat in a meal. Exclusive breast-feeding can cover the vitamin A needs of infants if the mother has an adequate vitamin A status and a large enough volume of milk. If the nursing mother has inadequate vitamin A intake or concomitant diseases or her infant was a preterm delivery, breast milk probably will not supply enough vitamin A to prevent deficiency. In developing countries, chronic dietary deficiency is the main cause of vitamin A deficiency and is exacerbated by infection. In early childhood, low vitamin A status results from inadequate intakes of animal food sources and edible oils, both of which are expensive, coupled with seasonal unavailability of vegetables and fruits and lack of marketed fortified food products. Factors that interfere with vitamin A metabolism may also affect status or function. For example, concurrent zinc deficiency can interfere with the mobilization of vitamin A from liver stores. Alcohol interferes with the conversion of retinol to retinaldehyde in the eye by competing for alcohol (retinol) dehydrogenase. Drugs that interfere with the absorption of vitamin A include mineral oil, neomycin, and bile acid sequestrants (e.g., cholestyramine).

2530 Deficiency Vitamin A deficiency is endemic in areas where diets are chronically poor, especially in southern Asia, sub-Saharan Africa, some parts of Latin America, and the western Pacific, including parts of China. Vitamin A status is usually assessed by measuring serum retinol (normal range, 1.05–3.50 µmol/L [30–100 µg/dL]) or via dose-response tests or tests of dark adaptation. To assure a correct biochemical assessment of vitamin A status, a simultaneous assessment of the inflammatory status is needed (in analogy to the assessment of iron status); not doing so may result in an overestimation of vitamin A deficiency. Correction factors to adjust the measured plasma vitamin A levels to account for the influence of C-reactive protein and α_1 -acid glycoprotein are available. Stable isotopic or invasive liver biopsy methods are available to estimate total-body stores of vitamin A. As judged by deficient serum retinol (<0.70 µmol/L [20 µg/dL]), vitamin A deficiency worldwide is present in 190 million preschool-age children, among whom >5 million have an ocular manifestation of deficiency termed *xerophthalmia*. This condition includes milder stages of night blindness and conjunctival *xerosis* (dryness) with *Bitot's spots* (white patches of keratinized epithelium appearing on the sclera) that may affect 1–5% of children in deficient populations as well as rare, potentially blinding corneal ulceration and necrosis. *Keratomalacia* (softening of the cornea) leads to corneal scarring that blinds an estimated quarter of a million children each year and is associated with fatality rates of 4–25%. However, vitamin A deficiency severe enough to cause any clinical stage poses an increased risk of death from diarrhea, dysentery, measles, malaria, or respiratory disease. This is because vitamin A deficiency can compromise barrier, innate, and acquired immune defenses to infection. In areas where deficiency is widely prevalent, vitamin A supplementation can markedly reduce the risk of childhood mortality (by 23–34%, on average). About 10% of pregnant women in undernourished settings also develop night blindness (assessed by history) during the latter half of pregnancy; this level of moderate to severe vitamin A deficiency is associated with an increased risk of maternal infection and death. Maternal vitamin A deficiency may also exacerbate already low vitamin A nutrition and associated risks for the newborn. In South Asia, where maternal deficiency is prominent, giving infants a single oral dose (50,000 IU) of vitamin A shortly after birth has reduced infant mortality by $\geq 10\%$, whereas in African settings less affected by maternal vitamin A deficiency, no effect has been noted, revealing differences in risk of deficiency and benefit of supplementation across regions. However, the World Health Organization does not recommend high-dose supplementation to newborns.

TREATMENT

Vitamin A Deficiency

Vitamin A is commercially available for treatment and prevention in esterified forms (e.g., acetate, palmitate), which are more stable than other forms. Any stage of xerophthalmia should be treated with 60 mg (or RAE) or 200,000 IU of vitamin A in oily solution, usually contained in a soft-gel capsule. The same dose is repeated 1 and 14 days later. Doses should be reduced by half for patients 6–11 months of age. Mothers with night blindness or Bitot's spots should be given vitamin A orally 3 mg daily for at least 3 months. These regimens are efficacious, and they are far less expensive and more widely available than injectable water-miscible vitamin A. A common approach to prevention is to provide vitamin A supplementation every 4–6 months to young children 6 months to 5 years of age (both HIV-positive and HIV-negative) in high-risk areas. For prevention, infants 6–11 months of age should receive 30 mg of vitamin A; children 12–59 months of age should receive 60 mg. For reasons that are not clear, although early neonatal vitamin A may reduce infant mortality, vitamin A given between 1 and 5 months of age has not proven effective in improving survival in high-risk settings.

Uncomplicated vitamin A deficiency is rare in industrialized countries. One high-risk group—extremely low-birth-weight (<1000 g) infants—is likely to be vitamin A deficient and should receive a

supplement of 1500 µg (or RAE) three times a week for 4 weeks. Severe measles in any society can lead to secondary vitamin A deficiency. Children hospitalized with measles should receive two 60-mg doses of vitamin A on 2 consecutive days. Vitamin A deficiency most often occurs in patients with malabsorptive diseases (e.g., celiac sprue, short-bowel syndrome) who have abnormal dark adaptation or symptoms of night blindness without other ocular changes. Typically, such patients are diagnosed in advanced care settings where they are treated for 1 month with 15 mg/d of a water-miscible preparation of vitamin A. This treatment is followed by a lower maintenance dose, with the exact amount determined by monitoring serum retinol. Finding application elsewhere in medicine, retinoic acid is useful in the treatment of promyelocytic leukemia (*Chap. 104*) and also is used in the treatment of cystic acne because it inhibits keratinization, decreases sebum secretion, and possibly alters the inflammatory reaction (*Chap. 57*).

No specific signs or symptoms result from carotenoid deficiency. It was postulated that β -carotene would be an effective chemopreventive agent for cancer because numerous epidemiologic studies had shown that diets high in β -carotene were associated with lower incidences of cancers of the respiratory and digestive systems. However, intervention studies in smokers found that treatment with high doses of β -carotene actually resulted in more lung cancers than did treatment with placebo. Non-provitamin A carotenoids such as lutein and zeaxanthin have been suggested to confer protection against macular degeneration, and one large-scale intervention study did not show a beneficial effect except in those with a low lutein status. The use of the non-provitamin A carotenoid lycopene to protect against prostate cancer has been proposed. However, the effectiveness of these agents has not been proved by intervention studies, and the mechanisms underlying these purported biologic actions are unknown.

Selective plant-breeding techniques that lead to a higher provitamin A carotenoid content in staple foods may decrease vitamin A malnutrition in low-income countries. Moreover, a recently developed genetically modified food (Golden Rice) had a β -carotene-to-vitamin A conversion ratio of ~3:1 in children.

Toxicity The acute toxicity of vitamin A was first noted in Arctic explorers who ate polar bear liver and has also been seen after administration of 150 mg to adults or 100 mg to children. Acute toxicity is manifested by increased intracranial pressure, vertigo, diplopia, bulging fontanels (in children), seizures, and exfoliative dermatitis; it may result in death. Among children being treated for vitamin A deficiency according to the protocols outlined above, transient bulging of fontanels occurs in 2% of infants, and transient nausea, vomiting, and headache occur in 5% of preschoolers. Chronic vitamin A intoxication is largely a concern in industrialized countries and has been seen in otherwise healthy adults who ingest 15 mg/d and children who ingest 6 mg/d over a period of several months. Manifestations include dry skin, cheilosis, glossitis, vomiting, alopecia, bone demineralization and pain, hypercalcemia, lymph node enlargement, hyperlipidemia, amenorrhea, and features of pseudotumor cerebri with increased intracranial pressure and papilledema. Liver fibrosis with portal hypertension may also result from chronic vitamin A intoxication. Provision of vitamin A in excess to pregnant women has resulted in spontaneous abortion and in congenital malformations, including craniofacial abnormalities and valvular heart disease. In pregnancy, the daily dose of vitamin A should not exceed 3 mg. Also, topical retinoids should be avoided during pregnancy. Commercially available retinoid derivatives are also toxic, including 13-cis-retinoic acid, which has been associated with birth defects. Thus, contraception should be continued for at least 1 year and possibly longer in women who have taken 13-cis-retinoic acid.

In malnourished children, vitamin A supplements (30–60 mg), in amounts calculated as a function of age and given in several rounds over 2 years, are considered to amplify nonspecific effects of vaccines. However, for unclear reasons, in one African setting, there has been a negative effect on mortality rates in incompletely vaccinated girls.

High doses of supplemental carotenoids do not result in toxic symptoms but should be avoided in smokers due to an increased risk of lung cancer. Very high doses of β -carotene (~200 mg/d) have been used to treat or prevent the skin rashes of erythropoietic protoporphyria. Carotenemia, which is characterized by a yellowing of the skin (in creases of the palms and soles) but not the sclerae, may follow ingestion of >30 mg of β -carotene daily. Hypothyroid patients are particularly susceptible to the development of carotenemia due to impaired breakdown of carotene to vitamin A. Reduction of carotenes in the diet results in the disappearance of skin yellowing and carotenemia over a period of 30–60 days.

VITAMIN D

The metabolism of the fat-soluble vitamin D is described in detail in [Chap. 409](#). The biologic effects of this vitamin are mediated by vitamin D receptors, which are found in most tissues; binding with these receptors potentially expands vitamin D actions to many different cell systems and organs (e.g., immune cells, brain, breast, colon, and prostate) in addition to the classic endocrine effects on calcium and phosphate metabolism and bone health. Vitamin D is thought to be important for maintaining normal function of many nonskeletal tissues such as muscle (including heart muscle), for immune function, and for inflammation as well as for cell proliferation and differentiation. Older studies have shown that vitamin D may be useful as adjunctive treatment for tuberculosis, psoriasis, and multiple sclerosis or for the prevention of certain cancers. Vitamin D insufficiency may increase the risk of type 1 diabetes mellitus, cardiovascular disease (insulin resistance, hypertension, or low-grade inflammation), or brain dysfunction (e.g., depression). However, the exact physiologic roles of vitamin D in these nonskeletal diseases and the importance of these roles have so far not been clarified. Recent placebo-controlled studies did not show a therapeutic benefit of vitamin D for cancer prevention, control of cardiovascular disease, or risk of type 2 diabetes, depression, tuberculosis infection, or other respiratory infections. Presently, it is not known whether these effects of vitamin D supplements (with or without calcium) might be different according to the baseline status (normal vs severely deficient) of patients.

The skin is a major source of vitamin D, which is synthesized upon skin exposure to ultraviolet B radiation (UV-B; wavelength, 290–320 nm). Except for fish, food (unless fortified) contains only limited amounts of vitamin D. Vitamin D₃ (ergocalciferol) is obtained from plant sources and is the chemical form found in some supplements.

Deficiency Vitamin D status has been assessed by measuring serum levels of 25-dihydroxyvitamin D (25(OH) vitamin D); however, there is no consensus on a uniform assay, on optimal serum levels, or on the real benefit of biochemical screening. The optimal level might, in fact, differ according to the targeted disease entity. Epidemiologic and experimental data indicate that a 25(OH) vitamin D level of >20 ng/mL (≥ 50 nmol/L; to convert ng/mL to nmol/L, multiply by 2.496) is sufficient for good bone health. The latter 25(OH) vitamin D plasma concentration would cover the requirements of 97.5% of the population. Some experts, however, advocate higher serum levels (e.g., >30 ng/mL) for other desirable endpoints of vitamin D action. There is insufficient evidence to recommend combined vitamin D and calcium supplementation as a primary preventive strategy (as opposed to secondary prevention) for reduction of the incidence of fractures in healthy men and premenopausal women.

Risk factors for vitamin D deficiency are old age, lack of sun exposure, dark skin (especially among residents of northern latitudes), fat malabsorption, and obesity; deficiency can also occur after gastric bypass surgery. In addition, in African populations, the prevalence of vitamin D deficiency might be high (especially in women, newborn babies, urban populations, and those living in northern African countries). *Rickets* represents the classic disease of vitamin D deficiency. Signs of deficiency are muscle soreness, weakness, and bone pain. Some of these effects are independent of calcium intake. To prevent glucocorticoid-induced osteoporosis, treatment with calcium (1000–1200 mg/d) and vitamin D (600–800 IU/d) through diet and/or supplements in combination with weight-bearing exercise is recommended.

The U.S. National Academy of Sciences recently advised that the majority of adult North Americans should receive 600 IU/d of vitamin D (RDA = 15 μ g/d or 600 IU/d; [Chap. 332](#)). However, for people aged >70 years, the RDA is set at 20 μ g/d (800 IU/d). The consumption of fortified or enriched foods as well as suberythemal sun exposure should be encouraged for people at risk for vitamin D deficiency. If adequate intake is impossible, vitamin D supplements should be taken, especially during the winter months. Vitamin D deficiency can be treated by oral administration of 50,000 IU/week for 6–8 weeks followed by a maintenance dose of 800 IU/d (20 μ g/d) from food and supplements once normal plasma levels have been attained. There is still uncertainty regarding the optimal therapeutic dosage (high vs low) for elderly at risk of falls. The physiologic effects of vitamin D₂ and vitamin D₃ are similar when these vitamins are ingested over long periods.

Toxicity The upper limit of intake has been set at 4000 IU/d. Contrary to earlier beliefs, acute vitamin D intoxication is rare and usually is caused by the uncontrolled and excessive ingestion of supplements or by faulty food fortification practices. High plasma levels of 1,25(OH)₂ vitamin D and calcium are central features of toxicity and mandate discontinuation of vitamin D and calcium supplements; in addition, treatment of hypercalcemia may be required.

VITAMIN E

Vitamin E is the collective designation for all stereoisomers of tocopherols and tocotrienols, although only the α -tocopherols meet human requirements. Vitamin E acts as a chain-breaking antioxidant and is an efficient peroxy radical scavenger that protects low-density lipoproteins and polyunsaturated fats in membranes from oxidation. A network of other antioxidants (e.g., vitamin C, glutathione) and enzymes maintains vitamin E in a reduced state. Vitamin E also inhibits prostaglandin synthesis and the activities of protein kinase C and phospholipase A₂.

Absorption and Metabolism After absorption, vitamin E is taken up from chylomicrons by the liver, and a hepatic α -tocopherol transport protein mediates intracellular vitamin E transport and incorporation into very-low-density lipoprotein. The transport protein has a particular affinity for the RRR isomeric form of α -tocopherol; thus, this natural isomer has the most biologic activity.

Requirement Vitamin E is widely distributed in the food supply, with particularly high levels in sunflower oil, safflower oil, and wheat germ oil; γ -tocotrienols are notably present in soybean and corn oils. Vitamin E is also found in meats, nuts, and cereal grains, and small amounts are present in fruits and vegetables. Vitamin E pills containing doses of 50–1000 mg are ingested by ~10% of the U.S. population. The RDA for vitamin E is 15 mg/d (34.9 μ mol or 22.5 IU) for all adults. Diets high in polyunsaturated fats may necessitate a slightly higher intake of vitamin E.

Dietary deficiency of vitamin E does not exist in developed countries but can occur in developing countries due to inadequate intake. Vitamin E deficiency is seen only in severe and prolonged malabsorptive diseases, such as celiac disease, chronic cholestatic liver disease, or after small-intestinal resection or bariatric surgery. Children with cystic fibrosis or prolonged cholestasis may develop vitamin E deficiency characterized by areflexia and hemolytic anemia. Children with abetalipoproteinemia cannot absorb or transport vitamin E and become deficient quite rapidly. A familial form of isolated vitamin E deficiency also exists; it is due to a defect in the α -tocopherol transport protein. Vitamin E deficiency causes axonal degeneration of the large myelinated axons and results in posterior column and spinocerebellar symptoms. Peripheral neuropathy is initially characterized by areflexia, with progression to an ataxic gait, and by decreased vibration and position sensations. Ophthalmoplegia, skeletal myopathy, and pigmented retinopathy may also be features of vitamin E deficiency. A deficiency of either vitamin E or selenium in the host has been shown to increase certain viral mutations and, therefore, virulence. The laboratory diagnosis of vitamin E deficiency is based on low blood levels of α -tocopherol (<5 μ g/mL, or <0.8 mg of α -tocopherol per gram of total lipids).

TREATMENT

Vitamin E Deficiency

Symptomatic vitamin E deficiency should be treated with 800–1200 mg of α -tocopherol per day. Patients with abetalipoproteinemia may need as much as 5000–7000 mg/d. Children with symptomatic vitamin E deficiency should be treated orally with water-miscible esters (400 mg/d); alternatively, 2 mg/kg per d may be administered intramuscularly. Vitamin E in high doses may protect against oxygen-induced retrolental fibroplasia and bronchopulmonary dysplasia as well as intraventricular hemorrhage of prematurity. Vitamin E has been suggested to increase sexual performance, treat intermittent claudication, and slow the aging process, but convincing evidence for these properties is lacking. When given in combination with other antioxidants, vitamin E may help prevent macular degeneration. Vitamin E may have favorable therapeutic effects in noncirrhotic nondiabetic patients with nonalcoholic steatohepatitis. High doses (60–800 mg/d) of vitamin E have been shown in controlled trials to improve parameters of immune function and reduce colds in nursing home residents, but intervention studies using vitamin E to prevent cardiovascular disease or cancer have not shown efficacy, and at doses >400 mg/d, vitamin E may even increase all-cause mortality rates and prostate cancer risk (especially in combination with selenium supplements).

Toxicity All forms of vitamin E are absorbed and could contribute to toxicity; however, the toxicity risk seems to be rather low as long as liver function is normal. High doses of vitamin E (>800 mg/d) may reduce platelet aggregation and interfere with vitamin K metabolism and are therefore contraindicated in patients taking warfarin and antiplatelet agents (such as aspirin or clopidogrel). Nausea, flatulence, and diarrhea have been reported at doses >1 g/d.

VITAMIN K

There are two natural forms of vitamin K: vitamin K₁, also known as *phylloquinone*, from vegetable sources, and vitamin K₂, or *menaquinones*, which are synthesized by bacterial flora and found in hepatic tissue. Phylloquinone can be converted to menaquinone in some organs.

Vitamin K is required for the posttranslational carboxylation of glutamic acid, which is necessary for calcium binding to γ -carboxylated proteins such as prothrombin (factor II); factors VII, IX, and X; protein C; protein S; and proteins found in bone (osteocalcin) and vascular smooth muscle (e.g., matrix Gla protein). However, the importance of vitamin K for bone mineralization and prevention of vascular calcification is not known. Warfarin-type drugs inhibit γ -carboxylation by preventing the conversion of vitamin K to its active hydroquinone form.

Dietary Sources Vitamin K is found in green leafy vegetables such as kale and spinach, and appreciable amounts are also present in margarine and liver. Vitamin K is present in vegetable oils; olive, canola, and soybean oils are particularly rich sources. The average daily intake by Americans is estimated to be ~100 µg/d.

Deficiency The symptoms of vitamin K deficiency are due to hemorrhage; newborns are particularly susceptible because of low fat stores, low breast milk levels of vitamin K, relative sterility of the infantile intestinal tract, liver immaturity, and poor placental transport. Intracranial bleeding as well as gastrointestinal and skin bleeding can occur in vitamin K-deficient infants 1–7 days after birth. Thus, vitamin K (0.5–1 mg IM) is given prophylactically at delivery.

Vitamin K deficiency in adults may be seen in patients with chronic small-intestinal disease (e.g., celiac disease, Crohn's disease), in those with obstructed biliary tracts, or after small-bowel resection. Broad-spectrum antibiotic treatment can precipitate vitamin K deficiency by reducing numbers of gut bacteria, which synthesize menaquinones, and by inhibiting the metabolism of vitamin K. In patients with warfarin therapy, the antidiabetes drug orlistat can lead to changes in international normalized ratio due to vitamin K

malabsorption. The assessment of the vitamin K status can be done by measurement of phylloquinone (vitamin K₁) concentration in serum (deficiency <0.15 µg/L); the cellular utilization of vitamin K can be assessed by the serum or plasma concentration of undercarboxylated prothrombin (protein induced by vitamin K absence/antagonism [PIVKA-II]). An elevated prothrombin time or activated partial thromboplastin time or reduced clotting factors are useful markers in severe deficiency but are otherwise nonspecific and lack sensitivity. Vitamin K deficiency is treated with a parenteral dose of 10 mg. For patients with chronic malabsorption, 1–2 mg/d should be given orally or 1–2 mg per week can be taken parenterally. Patients with liver disease may have an elevated prothrombin time because of liver cell destruction as well as vitamin K deficiency. If an elevated prothrombin time does not improve during vitamin K therapy, it can be deduced that this abnormality is not the result of vitamin K deficiency.

Toxicity Toxicity from dietary phylloquinones and menaquinones has not been described. High doses of vitamin K can impair the actions of oral vitamin K antagonist anticoagulants.

MINERALS

See also Table 333-2.

CALCIUM

See Chap. 409.

ZINC

Zinc is an integral component of many metalloenzymes in the body; it is involved in the synthesis and stabilization of proteins, DNA, and RNA and plays a structural role in ribosomes and membranes. Zinc is necessary for the binding of steroid hormone receptors and several other transcription factors to DNA. Zinc is essential for normal spermatogenesis, fetal growth, and embryonic development.

Absorption The absorption of zinc from the diet is inhibited by dietary phytate, fiber, oxalate, iron, and copper as well as by certain drugs, including penicillamine, sodium valproate, and ethambutol. Protein-containing foods, i.e., meat, shellfish, nuts, and legumes, are good sources of bioavailable zinc, whereas zinc in grains and legumes is less available for absorption. Grains and legumes contain phytate that binds zinc in the intestine and reduces its availability for absorption.

Deficiency Mild zinc deficiency has been described in many diseases, including diabetes mellitus, HIV/AIDS, cirrhosis, alcoholism, inflammatory bowel disease, malabsorption syndromes, and sickle cell disease. In these diseases, mild chronic zinc deficiency can cause stunted growth in children, decreased taste sensation (*hypogeusia*), and impaired immune function. Severe chronic zinc deficiency has been described as a cause of hypogonadism and dwarfism in several Middle Eastern countries. In these children, hypopigmented hair is also part of the syndrome. Acrodermatitis enteropathica is a rare autosomal recessive disorder characterized by abnormalities in zinc absorption. Clinical manifestations include diarrhea, alopecia, muscle wasting, depression, irritability, and a rash involving the extremities, face, and perineum. The rash is characterized by vesicular and pustular crusting with scaling and erythema. Occasional patients with Wilson's disease have developed zinc deficiency as a consequence of penicillamine therapy (Chap. 415).

Zinc deficiency is prevalent in many developing countries and usually coexists with other micronutrient deficiencies (especially iron deficiency). Zinc (20 mg/d until recovery) may be an effective adjunctive therapeutic strategy for diarrheal disease and pneumonia in children \leq 6 months of age.

The diagnosis of zinc deficiency is usually based on a serum zinc level <12 µmol/L (<70 µg/dL). Pregnancy and birth control pills may cause a slight depression in serum zinc levels, and hypoalbuminemia from any cause can result in hypozincemia. In acute stress situations (illness, but also postexercise recovery), zinc may be redistributed from serum into tissues. Zinc deficiency may be treated with 60 mg of elemental zinc taken by mouth twice a day. Zinc gluconate lozenges (13 mg of elemental zinc every 2 h while awake) have been reported to

TABLE 333-2 Deficiencies and Toxicities of Metals

ELEMENT	DEFICIENCY	TOXICITY	TOLERABLE UPPER (DIETARY) INTAKE LEVEL
Boron	No biologic function determined	Developmental defects, male sterility, testicular atrophy	20 mg/d (extrapolated from animal data)
Calcium	Reduced bone mass, osteoporosis	Renal insufficiency (milk-alkali syndrome), nephrolithiasis, impaired iron absorption, thiazide diuretics	2500 mg/d (milk-alkali)
Copper	Anemia, growth retardation, defective keratinization and pigmentation of hair, hypothermia, degenerative changes in aortic elastin, osteopenia, intellectual disability	Nausea, vomiting, diarrhea, hepatic failure, tremor, psychiatric disturbances, hemolytic anemia, renal dysfunction	10 mg/d (liver toxicity)
Chromium	Impaired glucose tolerance	<i>Occupational:</i> Renal failure, dermatitis, pulmonary cancer	Not determined
Fluoride	↑ Dental caries	Dental and skeletal fluorosis, osteosclerosis	10 mg/d (fluorosis)
Iodine	Thyroid enlargement, ↓ T ₄ , cretinism	Thyroid dysfunction, acne-like eruptions	1100 µg/d (thyroid dysfunction)
Iron	Muscle abnormalities, koilonychia, pica, anemia, ↓ work performance, impaired cognitive development, premature labor, ↑ perinatal maternal death	Gastrointestinal effects (nausea, vomiting, diarrhea, constipation), iron overload with organ damage, acute and chronic systemic toxicity, increased susceptibility to malaria, increased risk association with certain chronic diseases (e.g., diabetes)	45 mg/d of elemental iron (gastrointestinal side effects)
Manganese	Impaired growth and skeletal development, reproduction, lipid and carbohydrate metabolism; upper body rash	<i>General:</i> Neurotoxicity, Parkinson-like symptoms <i>Occupational:</i> Encephalitis-like syndrome, Parkinson-like syndrome, psychosis, pneumoconiosis	11 mg/d (neurotoxicity)
Molybdenum	Severe neurologic abnormalities	Reproductive and fetal abnormalities	2 mg/d (extrapolated from animal data)
Selenium	Cardiomyopathy, heart failure, striated muscle degeneration	<i>General:</i> Alopecia, nausea, vomiting, abnormal nails, emotional lability, peripheral neuropathy, lassitude, garlic odor to breath, dermatitis <i>Occupational:</i> Lung and nasal carcinomas, liver necrosis, pulmonary inflammation	400 µg/d (hair, nail changes)
Phosphorus	Rickets (osteomalacia), proximal muscle weakness, rhabdomyolysis, paresthesia, ataxia, seizure, confusion, heart failure, hemolysis, acidosis	Hyperphosphatemia	4000 mg/d
Zinc	Growth retardation, ↓ taste and smell, alopecia, dermatitis, diarrhea, immune dysfunction, failure to thrive, gonadal atrophy, congenital malformations	<i>General:</i> Reduced copper absorption, gastritis, sweating, fever, nausea, vomiting <i>Occupational:</i> Respiratory distress, pulmonary fibrosis	40 mg/d (impaired copper metabolism)

reduce the duration and symptoms of the common cold in adults, but study results are conflicting.

Toxicity Acute zinc toxicity after oral ingestion causes nausea, vomiting, and fever. Zinc fumes from welding may also be toxic and cause fever, respiratory distress, excessive salivation, sweating, and headache. Chronic large doses of zinc (ranging from 150 to 450 mg/d) may depress immune function and cause hypochromic anemia as a result of a secondary copper deficiency. Intranasal zinc preparations should be avoided because they may lead to irreversible damage of the nasal mucosa and anosmia.

■ COPPER

Copper is an integral part of numerous enzyme systems, including amine oxidases, ferroxidase (ceruloplasmin), cytochrome c oxidase, superoxide dismutase, and dopamine hydroxylase. Copper is also a component of ferroprotein, a transport protein involved in the basolateral transfer of iron during absorption from the enterocyte. As such, copper plays a role in iron metabolism, melanin synthesis, energy production, neurotransmitter synthesis, and CNS function; the synthesis and cross-linking of elastin and collagen; and the scavenging of superoxide radicals. Dietary sources of copper include shellfish, liver, nuts, legumes, bran, and organ meats.

Deficiency Dietary copper deficiency is relatively rare, although it has been described in premature infants who are fed milk diets and in infants with malabsorption (Table 333-2). Copper-deficiency anemia (refractory to therapeutic iron) has been reported in patients with malabsorptive diseases and nephrotic syndrome and in patients treated for Wilson's disease with chronic high doses of oral zinc, which can interfere with copper absorption. *Menkes kinky hair syndrome* is an

X-linked metabolic disturbance of copper metabolism characterized by intellectual disability, hypocupremia, and decreased circulating ceruloplasmin (Chap. 413). This syndrome is caused by mutations in the copper-transporting ATP7A gene. Children with this disease often die within 5 years because of dissecting aneurysms or cardiac rupture. Aceruloplasminemia is a rare autosomal recessive disease characterized by tissue iron overload, mental deterioration, microcytic anemia, and low serum iron and copper concentrations.

The diagnosis of copper deficiency is usually based on low serum levels of copper (<65 µg/dL) and low ceruloplasmin levels (<20 mg/dL). Serum levels of copper may be elevated in pregnancy or stress conditions since ceruloplasmin is an acute-phase reactant and 90% of circulating copper is bound to ceruloplasmin. It has been suggested that mild or subclinical copper deficiency is more common than expected; at-risk individuals include patients with cholestasis or chronic diarrheal diseases, dialysis patients, and people on long-term zinc supplements. The role of copper in cardiovascular disease, immune function, bone health, or neurodegenerative diseases is still unclear.

Toxicity Copper toxicity is usually accidental (Table 333-2). In severe cases, kidney failure, liver failure, and coma may ensue. In Wilson's disease, mutations in the copper-transporting ATP7B gene lead to accumulation of copper in the liver and brain, with low blood levels due to decreased ceruloplasmin (Chap. 415). A potential negative role of copper in the pathogenesis of Alzheimer's disease has been reported.

■ SELENIUM

Selenium, in the form of selenocysteine, is a component of the enzyme glutathione peroxidase, which serves to protect proteins, cell membranes, lipids, and nucleic acids from oxidant molecules. As such, selenium is being actively studied as a chemopreventive agent against

certain cancers, such as prostate cancer. However, it remains unclear whether selenium is effective as a chemopreventive agent or whether it increases cancer risk (e.g., prostate cancer). Convincing evidence for a protective effect of selenium on cognitive decline or cardiovascular disease risk is presently lacking. Selenocysteine is also found in the deiodinase enzymes, which mediate the deiodination of thyroxine to triiodothyronine (Chap. 382). Rich dietary sources of selenium include seafood, muscle meat, and cereals, although the selenium content of cereal is determined by the soil concentration. Countries with low soil concentrations include parts of Scandinavia, China, and New Zealand. *Keshan disease* is an endemic cardiomyopathy found in children and young women residing in regions of China where dietary intake of selenium is low (<20 µg/d). Concomitant deficiencies of iodine and selenium may worsen the clinical manifestations of cretinism. Chronic ingestion of large amounts of selenium leads to selenosis, characterized by hair and nail brittleness and loss, garlic breath odor, skin rash, myopathy, irritability, and other abnormalities of the nervous system.

■ CHROMIUM

Chromium potentiates the action of insulin in patients with impaired glucose tolerance, presumably by increasing insulin receptor-mediated signaling, although its usefulness in treating type 2 diabetes is uncertain. In addition, improvement in blood lipid profiles has been reported in some patients. The usefulness of chromium supplements in muscle building has not been substantiated. Rich food sources of chromium include yeast, meat, and grain products. Chromium in the trivalent state is found in supplements and is largely nontoxic; however, chromium-6 is a product of stainless steel welding and is a known pulmonary carcinogen as well as a cause of liver, kidney, and CNS damage.

■ MAGNESIUM

See Chap. 409.

■ FLUORIDE, MANGANESE, AND ULTRATRACE ELEMENTS

An essential function for fluoride in humans has not been described, although it is useful for the maintenance of structure in teeth and bones. Adult fluorosis results in mottled and pitted defects in tooth enamel as well as brittle bone (skeletal fluorosis).

Manganese and molybdenum deficiencies have been reported in patients with rare genetic abnormalities and in a few patients receiving prolonged total parenteral nutrition. Several manganese-specific enzymes have been identified (e.g., manganese superoxide dismutase). Deficiencies of manganese have been reported to result in bone demineralization, poor growth, ataxia, disturbances in carbohydrate and lipid metabolism, and convulsions.

Ultratrace elements are defined as those needed in amounts <1 mg/d. Essentiality has not been established for most ultratrace elements, although selenium, chromium, and iodine are clearly essential (Chap. 382). Molybdenum is necessary for the activity of sulfite and xanthine oxidase, and molybdenum deficiency may result in skeletal and brain lesions.

■ FURTHER READING

- C GF J , M JP: *The Vitamins: Fundamental Aspects in Nutrition and Health*. 5th ed. London, Academic Press, 2017, p 612.
- I A et al: Vitamin A supplementation for preventing morbidity and mortality in children from six months to five years of age. Cochrane Database Syst Rev 3:CD008524, 2017.
- L ZS et al: Zinc supplementation for the promotion of growth and prevention of infections in infants less than six months of age. Cochrane Database Syst Rev 4:CD010205, 2020.
- M JL et al: Clinical practice guidelines for the perioperative nutrition, metabolic, and nonsurgical support of patients undergoing bariatric procedures—2019 update. Surg Obes Relat Dis 16:175, 2020.
- N SM et al: Methodologic approach for the Biomarkers Reflecting Inflammation and Nutritional Determinants of Anemia (BRINDA) project. Am J Clin Nutr 106(Suppl 1):333S, 2017.
- N B et al: Targeting cancer vulnerabilities with high-dose vitamin C. Nat Rev Cancer 19:271, 2020.

O Y et al: Comprehensive review of Wernicke encephalopathy: Pathophysiology, clinical symptoms and imaging findings. Jpn J Radiol 38:809, 2020.

S GA et al: Trends and mortality effects of vitamin A deficiency in children in 138 low-income and middle-income countries between 1991 and 2013: A pooled analysis of population-based surveys. Lancet Glob Health 3:e528, 2015.

T SA et al: Biomarkers of nutrition for development (BOND): Vitamin A review. J Nutr 146:1816S, 2016.

V M, R KJ: More results but no clear conclusion on selenium and cancer. Am J Clin Nutr 104:245, 2016.

W H O : Guideline: Vitamin A supplementation in pregnant women. Geneva, World Health Organization, 2011.

334

Malnutrition and Nutritional Assessment

Gordon L. Jensen



Malnutrition occurs in 30–50% of hospitalized patients depending on the setting and criteria that are used. Poor wound healing, compromised immune status, impaired organ function, increased length of hospital stay, and increased mortality are among the notable adverse outcomes associated with malnutrition. It is now widely appreciated that acute or chronic inflammation contribute to the pathophysiology of disease-related or injury-related malnutrition. The presence of inflammation can also render historic nutrition assessment indicators, like albumin and prealbumin, unreliable, and inflammation diminishes favorable responses to nutrition therapies. In order to guide appropriate care, it is necessary to properly assess and diagnose malnutrition. Nutrition assessment is a comprehensive evaluation to diagnose a malnutrition syndrome and to guide intervention and expected outcomes. Patients are often targeted for assessment after being identified at nutritional risk based on screening procedures conducted by nursing or nutrition personnel within 24 h of hospital admission. Screening tends to focus explicitly on a few risk variables like weight loss, compromised dietary intake, and high-risk medical/surgical diagnoses. Preferably, health professionals complement this screening with a systematic approach to comprehensive nutrition assessment that incorporates an appreciation for the contributions of inflammation that serve as the basis for new approaches to the diagnosis and management of malnutrition syndromes.

■ MALNUTRITION SYNDROMES

Famine and starvation have long been leading causes of malnutrition and remain so in developing countries. However, with improvements in agriculture, education, public health, health care, and living standards, malnutrition in the settings of disease, surgery, and injury has become a prevalent concern throughout the world. Malnutrition now encompasses the full continuum of undernutrition and overnutrition (obesity). For the objectives of this chapter, we will focus upon the former. Historic definitions for malnutrition syndromes are problematic in their use of diagnostic criteria that suffer poor sensitivity, sensitivity, and interobserver reliability. Definitions overlap, and confusion and misdiagnosis are frequent. In addition, some approaches do not recognize undernutrition in obese persons. While the historic syndromes of marasmus, kwashiorkor, and protein-calorie malnutrition remain in use, this chapter will instead highlight evolving insights to the diagnosis of malnutrition syndromes.

The Subjective Global Assessment, a comprehensive nutrition assessment that included a metabolic stress of disease component, was described and validated in the 1980s. In 2010, an International Consensus Guideline Committee incorporated a new appreciation for

the role of inflammatory response into their proposed nomenclature for nutrition diagnosis in adults in the clinical practice setting. *Starvation-associated malnutrition* is when there is chronic starvation without inflammation, *chronic disease-associated malnutrition* is when inflammation is chronic and of mild to moderate degree, and *acute disease- or injury-associated malnutrition* is when inflammation is acute and of severe degree (see Table 334-1 for examples). In 2012, the Academy of Nutrition and Dietetics and the American Society for Parenteral

and Enteral Nutrition (ASPEN) extended this approach using clinical characteristics to support diagnosis, including the presence of illness or injury, poor food intake, weight loss, and physical findings of fat loss, muscle loss, edema, or reduced grip strength. In 2016, the European Society for Parenteral and Enteral Nutrition (ESPEN) formally adopted a disease/inflammation-based construct similar to these earlier approaches. Also in 2016, the Global Leadership Initiative on Malnutrition (GLIM), a collaborative effort of ASPEN, ESPEN, the

TABLE 334-1 History and Physical Examination Elements

ELEMENT	NOTES
Historical Data	
Body weight	Ask about usual weight, peak weight, and deliberate weight loss. A 4.5-kg (10-lb) weight loss over 6 months is noteworthy, and a weight loss of >10% of usual body weight is prognostic of clinical outcomes. Use medical records, family, and caregivers as information resources.
Medical and surgical conditions; chronic disease	<p>Look for medical or surgical conditions or chronic disease that can place one at nutritional risk secondary to increased requirements or compromised intake or assimilation such as: critical illness, severe burns, major abdominal surgery, multitrauma, closed head injury, previous gastrointestinal surgery, severe gastrointestinal hemorrhage, enterocutaneous fistula, gastrointestinal obstruction, mesenteric ischemia, severe acute pancreatitis, chronic pancreatitis, inflammatory bowel disease, celiac disease, bacterial overgrowth, solid or hematologic malignancy, bone marrow transplant, acquired immune deficiency syndrome, and organ failure/transplant—kidney, liver, heart, lung, or gut.</p> <p>A number of conditions or diseases are characterized by severe acute inflammatory response including critical illness, major infection/sepsis, adult respiratory distress syndrome, systemic inflammatory response syndrome, severe burns, major abdominal surgery, multitrauma, and closed head injury.</p> <p>Many conditions or diseases are more typically associated with mild to moderate chronic inflammatory response. Examples include cardiovascular disease, congestive heart failure, cystic fibrosis, inflammatory bowel disease, celiac disease, chronic pancreatitis, rheumatoid arthritis, solid tumors, hematologic malignancies, sarcopenic obesity, diabetes mellitus, metabolic syndrome, cerebrovascular accident, neuromuscular disease, dementia, organ failure/transplant (kidney, liver, heart, lung, or gut), periodontal disease, pressure wounds, and chronic obstructive pulmonary disease. Note that acute exacerbations, infections, or other complications may superimpose acute inflammatory response on such conditions or diseases.</p> <p>Examples of starvation-associated conditions that generally have little or no discernable inflammatory component include anorexia nervosa or compromised intake in the setting of major depression.</p>
Constitutional signs/ symptoms	Fever or hypothermia can indicate active inflammatory response. Tachycardia is also common. Anorexia is another manifestation of inflammatory response and is also often a side effect of treatments and medications.
Eating difficulties/ gastrointestinal complaints	Poor dentition or problems swallowing can compromise oral intake. Vomiting, nausea, abdominal pain, abdominal distension, diarrhea, constipation, and gastrointestinal bleeding can be signs of gastrointestinal pathology that may place one at nutritional risk.
Eating disorders	Look for distorted body image, compulsive exercise, amenorrhea, vomiting, tooth loss, dental caries, and use of laxatives, diuretics, or ipecac.
Medication use	Many medications can adversely affect nutrient intake or assimilation. Review potential drug-drug and drug-nutrient interactions. A pharmacist consultant can be helpful.
Dietary practices and supplement use	Look for dietary practices including therapeutic, weight reduction, vegetarian, macrobiotic, and fad diets. Also record use of dietary supplements, including vitamins, minerals, and herbs. Ask about dietary intake. Recall, record, and food frequency tools are available. It is estimated that 50% or more of adults take dietary supplements.
Influences on nutritional status	Ask about factors such as living environment, functional status (activities of daily living and instrumental activities of daily living), dependency, caregiver status, resources, dentition, alcohol or substance abuse, mental health (depression or dementia), and lifestyle.
Physical Examination Data	
Body mass index (BMI)	<p>BMI = weight in kg/(height in meters)²</p> <p>BMI <18.5 kg/m² proposed screen for malnutrition per National Institutes of Health guidelines. BMI ≤15 kg/m² or less is associated with increased mortality.</p> <p>Comparison with ideal body weight for stature can also be determined from reference tables. Note hydration status and edema at the time body weight is determined.</p>
Weight loss	<p>Look for loss of muscle mass and subcutaneous fat.</p> <p>Temporal and neck muscle wasting may be readily observed. Anthropometrics including skinfolds and circumferences can be useful but require training to achieve reliability.</p>
Weakness/loss of strength	Decreased handgrip and leg extensor strength have been related to loss of muscle mass in malnourished states. Lower extremity weakness may be observed in thiamine deficiency.
Peripheral edema	Peripheral edema may confound weight measurements and is often observed with reduced visceral proteins as well as inflammatory states. Edema may also be observed with thiamine deficiency.
Hair examination	<p>Hair findings are indicative of certain nutrient deficiencies.</p> <p>Loss: protein, vitamin B₁₂, folate</p> <p>Brittle: biotin</p> <p>Color change: zinc</p> <p>Dry: vitamins A and E</p> <p>Easy pluckability: protein, biotin, zinc</p> <p>Coiled, corkscrew: vitamins A and C</p> <p>Alopecia is common in severely malnourished persons.</p> <p>Ask about excessive hair loss on pillow or when combing hair.</p>

(Continued)

TABLE 334-1 History and Physical Examination Elements (Continued)

ELEMENT	NOTES
Skin examination	Skin findings are indicative of certain nutrient deficiencies. Desquamation: riboflavin Petechiae: vitamins A and C Perifollicular hemorrhage: vitamin C Ecchymosis: vitamins C and K Xerosis, bran-like desquamation: essential fatty acid Pigmentation, cracking, crusting: niacin Acneiform lesions, follicular keratosis, xerosis: vitamin A Acro-orificial dermatitis, erythematous, vesiculobullous, and pustular: zinc Characteristic nutritional dermatitis and skin findings may be observed with a number of nutrient deficiencies. Wounds and pressure sores should also be noted as indicators of compromised nutritional status.
Eye examination	Ocular findings are indicative of certain nutrient deficiencies. Bitot's spots: vitamin A Xerosis: vitamin A Angular palpebitis: riboflavin Also ask about difficulties with night vision/night blindness; indicates vitamin A deficiency.
Perioral examination	Perioral findings are indicative of certain nutrient deficiencies. Angular stomatitis and cheilosis: B complex, iron, protein Glossitis: niacin, folate, vitamin B ₁₂ Magenta tongue: riboflavin Bleeding gums, gingivitis, tooth loss: vitamin C Angular stomatitis, cheilosis, and glossitis are associated with vitamin and mineral deficiencies. Note poor dentition, caries, and tooth loss. Difficulty swallowing and impairment of gag should also be recognized.
Extremity examination	Extremity findings indicate certain nutrient deficiencies Arthralgia: vitamin C Calf pain: thiamine Extremities may also exhibit loss of muscle mass and/or peripheral edema. Neurologic findings in the extremities may also result from deficiencies described below.
Mental status/nervous system examination	Mental and nervous system findings indicate certain nutrient deficiencies. Ophthalmoplegia and foot drop: thiamine Paresthesia: thiamine, vitamin B ₁₂ , biotin Depressed vibratory and position senses: vitamin B ₁₂ Anxiety, depression, and hallucinations: niacin Memory disturbance: vitamin B ₁₂ Hyporeflexia, loss of lower extremity deep tendon reflexes: thiamine, vitamin B ₁₂ Conduct formal cognitive and depression assessments as appropriate. Dementia and depression are common causes of malnutrition among older persons. Wernicke-Korsakoff syndrome may be observed with severe thiamine deficiency.
Functional assessment	Observe and test physical performance as indicated: gait, chair stands, stair steps, and balance. These provide complex measures of integrated neurologic status, coordination, and strength.

Source: Reproduced with permission from GL Jensen: *Nutritional Syndromes*, In: Korenstein, D (Ed). ACP Smart Medicine [publisher archive]. Philadelphia (PA): American College of Physicians, 2013.

Latin American Federation of Parenteral and Enteral Nutrition, the Parenteral and Enteral Society of Asia, and other nutrition societies, embarked on an effort to build global consensus around commonly used evidence-based criteria for diagnosis of malnutrition in adults in clinical settings. Weight loss, low body mass index, and reduced muscle mass were selected as phenotypic criteria, whereas reduced food intake and disease burden/inflammation were selected as etiologic criteria. One phenotypic criterion and one etiologic criterion were deemed necessary for the preliminary diagnosis of malnutrition. Where available, this diagnosis should trigger comprehensive nutrition assessment by a skilled nutrition professional. However, the primary objective is to offer a simple approach that can be readily used in global settings with limited clinical nutrition resources. Recent studies suggest that these newer approaches to diagnosis of malnutrition have similar utility in predicting adverse outcomes. This is not surprising since they share a number of common criteria including a metabolic stress of disease component that is a proxy indicator of inflammation. Irrespective of the approach that is selected, assessment of patients can be facilitated using the indicators of malnutrition and inflammation described below.

■ NUTRITION ASSESSMENT

There is unfortunately no single clinical or laboratory indicator of comprehensive nutritional status. Assessment therefore requires systematic integration of data from a variety of sources. Micronutrient deficiencies of clinical relevance may be detected in association with any of the malnutrition syndromes, but a detailed discussion of their assessment is beyond the scope of this chapter (see Chap. 333). Physical findings characteristic of micronutrient deficiencies are, however, summarized in Table 334-1.

Medical/Surgical History and Clinical Diagnosis Knowledge of a patient's medical/surgical history and associated clinical diagnoses is especially helpful in discerning the likelihood of malnutrition and inflammation. Nonvolitional weight loss is a well-validated nutrition assessment indicator and is often also associated with underlying disease or inflammatory condition. The degree and duration of weight loss determine its clinical significance. A 10% loss of body weight over 6 months is of clinical relevance, whereas a 30% loss of body weight over the same duration is severe and life-threatening. Since weight loss history is often unavailable or unreliable, one should query the patient

as well as the medical records, family, and caregivers as appropriate to secure a valid weight trajectory.

A number of conditions or diseases are characterized by severe acute inflammatory response, whereas others are more typically associated with a chronic inflammatory response that is mild to moderate in severity and may be relapsing and remitting (Table 334-1). It is also common for acute inflammatory events to be superimposed on those with chronic conditions; for example, a patient with chronic renal disease is admitted to the hospital with sepsis. The inflammatory milieu, especially when severe, may modify nutrient requirements by elevating resting energy expenditure and promoting muscle catabolism and nitrogen losses. Inflammation also promotes anorexia, decreasing food intake and further compromising nutritional status. A deteriorating course may result because the presence of inflammation may reduce the benefit of nutritional interventions and the associated malnutrition may in turn diminish the effectiveness of medical therapies. It is also imperative to recognize medical/surgical conditions or diseases that place patients at increased risk to become malnourished because they have increased nutritional requirements or compromised intake or assimilation (Table 334-1).

Nutrition assessment should also include a review of medications with attention to undesirable side effects including anorexia, xerostomia, nausea, diarrhea, and constipation. Potential drug–nutrient interactions should also be identified.

Clinical Signs and Physical Examination Nonspecific clinical indicators of inflammation include fever, hypothermia, and tachycardia. The nutrition-focused physical examination should identify edema as well as signs of weight gain/loss and specific nutrient deficiencies. Thorough examination should be particularly directed to those parts of the body where high cell turnover occurs (e.g., hair, skin, mouth, tongue) as they are most likely to exhibit observable signs of nutritional deficiencies (Table 334-1). Physical findings of weight loss associated with decreased muscle and subcutaneous fat mass should not be overlooked, but when appreciable edema is present, these changes may not be readily appreciated.

Anthropometric Data Body weight measurements are recommended with each clinic visit or hospitalization so that a reliable weight change trajectory may be monitored. Patients should be weighed in a consistent manner without overgarments or shoes. In order to secure valid measurements, calibration of scales and appropriate staff training are essential. Chair or bed scales may be used for those who cannot stand. For those who are able, height should be measured in a standing position, without shoes, using a stadiometer. If an adult cannot safely stand, height can be estimated by doubling the arm span measurement (from the patient's sternal notch to the end of the longest finger). Stature of frail older persons can also be estimated from measurement of knee height using a caliper device.

Body weight is often standardized for height to obtain an ideal weight for comparison, but available reference tables require subjective assessment of frame size and offer limited reference data for many relevant population groups, including older persons. A simple measure of body size and an indirect measure of body fatness is provided by body mass index (BMI), defined as weight (kg)/height (m^2). The National Institutes of Health BMI categories for adults are as follows: BMI <18.5 = underweight, BMI 18.5–24.9 = desirable, BMI 25.0–29.9 = overweight, and BMI ≥ 30 = obese. Note that being overweight or obese does not mean that one cannot be severely malnourished due to inadequate nutrition intake or assimilation. Underweight status is not required for the diagnosis of malnutrition. While classical anthropometric measurements including skinfolds and circumferences can be helpful, their utility in routine patient care has been limited because practitioner training is required to achieve suitable reliability. Body composition assessment methodologies include bioelectrical impedance analysis (BIA), dual-energy x-ray absorptiometry (DEXA), computed tomography (CT), and magnetic resonance imaging (MRI). The imaging modalities have become the state of the art for precise measurements of muscle mass. It is possible to take advantage of CT or

MRI studies that are being done for other clinical purposes to evaluate musculature.

Laboratory Indicators Laboratory findings (Table 334-2) are but one part of the comprehensive nutrition assessment and must be used in combination with other domains of assessment to appropriately diagnose a malnutrition syndrome. Although serum albumin and pre-albumin are often measured in patients with suspected malnutrition, their utility is limited due to their poor sensitivity and specificity as indicators of nutritional status. Patients with low albumin or prealbumin may or may not prove to be malnourished when evaluated by comprehensive nutrition assessment because these proteins are readily reduced by the systemic response to injury, disease, or inflammation. C-reactive protein is a positive acute-phase reactant that may be measured to help discern whether active inflammation is manifest. If C-reactive protein is increased and albumin or prealbumin decreased, inflammation is likely to be a contributing factor. Since it is recognized that C-reactive protein suffers limitations as a point-in-time measure, trends in levels over the clinical course may be helpful. Research suggests that interleukin 6, and perhaps other cytokines, may also offer promise as indicators of inflammatory status. Nonspecific laboratory indicators that are often associated with inflammatory response include leukocytosis and hyperglycemia. Additional tests that may be obtained to help confirm the presence of inflammatory response include 24-h urine urea nitrogen and indirect calorimetry. In the setting of severe acute systemic inflammatory response, negative nitrogen balance and elevated resting energy expenditure are anticipated.

Dietary Assessment Dietary assessment can be used to detect inadequate or imbalanced food or nutrient intakes. While dietary assessment in patient care settings can be quite challenging, 24-h recall and modified diet history approaches are sometimes used. A modified diet history is targeted to query types and frequencies of intake of specific foods of interest. It is often necessary to access diverse resources for diet history information including the patient, medical records, family, and caregivers. Consultation of a registered dietitian nutritionist is highly recommended. Dietary practices and supplements should be carefully reviewed for potential inadequacies and toxicities. Since patients will often present to health care practitioners with acute medical events superimposed upon chronic health conditions, it is common for patients to have had decreased food intakes and malnutrition for extended periods prior to assessment. It is therefore imperative that compromised dietary intake should not be overlooked so that appropriate intervention may be undertaken.

Ongoing assessment is indicated when parenteral or enteral feedings are initiated, because it is necessary to discern what amount of formula is actually being administered to and received by the patient. Enteral feedings, in particular, are often interrupted or held for procedures, tolerance issues, and feeding tube displacements. It is therefore not unusual for such patients to be appreciably underfed for extended periods. When a patient is beginning to transition to oral feedings, it is imperative to monitor quantities of food and/or supplements that are actually consumed as well as patient tolerance to feeding. Meals are often delayed or missed for tests or procedures. If possible, the patient should be queried about intake since tray inspection is notoriously unreliable as an indicator of consumption.

Functional Outcomes Advanced malnutrition is accompanied by declines in muscle mass and function that can be detected by strength and physical performance measures. Handgrip strength measured with a simple handgrip dynamometer is the most practical routine clinical assessment. Physical performance tests such as timed gait, chair stands, and stair steps are used in the comprehensive assessment of integrated functions in frail older persons.

The decline in overall functional status observed in advanced malnutrition is associated with nutrient deficiencies and impairment of organ system functions. Poor wound healing and immune compromise are examples of such impairments. Improved wound healing parameters and restored responsiveness to recall antigens by delayed

TABLE 334-2 Body Composition, Laboratories, and Other Studies

TEST	NOTES
Body Composition Studies	
Anthropometrics	Skinfolds and circumferences require training for reliability. Typical coefficient of variation is $\geq 10\%$.
Bioelectrical impedance	Based upon differential resistance of body tissues. Equipment easily portable. Good measure of body water. Requires population-specific validation of regression equations.
Water displacement	Impractical for most clinical settings. Weighed in water tank. Historic reference measure.
Whole-body counting and isotope dilution techniques	Research methodologies. Naturally occurring ^{40}K isotope to measure body cell mass by whole-body counting. Total-body water measurement by dilution volume of tritium, deuterium, or ^{18}O -labeled water.
Air plethysmography	Subject sits inside moderately sized BodPod chamber. Validated against water displacement and impedance.
Dual energy x-ray absorptiometry (DEXA)	Often used for bone density but can be used for soft tissue measurements with appropriate software. Can compare truncal and appendicular components. Modest x-ray exposure.
Imaging with computed tomography (CT) or magnetic resonance imaging (MRI)	State of the art research methods for visualizing body tissue compartments. Can quantify visceral fat. Costly, and CT entails x-ray exposure.
Laboratories and Other Studies	
Albumin	Lacks sensitivity and specificity for malnutrition. Potent risk indicator for morbidity and mortality. Proxy measure for underlying injury, disease, or inflammation. Half-life is 14–20 days. Also consider liver disease, nephrotic syndrome, and protein-wasting enteropathy.
Prealbumin	Sensitive to short-term changes in inflammation and protein nutrition with half-life of 2–3 days. Otherwise suffers the same limitations of albumin with limited sensitivity and specificity for malnutrition. Levels may be decreased in liver failure and increased in renal failure.
Transferrin	Acute-phase reactant also altered by perturbation in iron status. Half-life is 8–10 days. Lacks sensitivity and specificity for malnutrition.
Retinol-binding protein	Responds to very-short-term changes in nutritional status, but utility is also limited by response to stress and inflammation. Half-life is 12 h. Also affected by vitamin A deficiency and renal disease.
C-reactive protein	C-reactive protein is a positive acute-phase reactant. It is generally elevated if an active inflammatory process is manifest.
Cholesterol	Low cholesterol ($<160\text{ mg/dL}$) is often observed in malnourished persons with serious underlying disease. It is unrelated to dietary intake in many clinical settings. Increased complications and mortality are observed. It appears that low cholesterol is again a nonspecific feature of poor health status that reflects cytokine-mediated inflammatory condition. Vegans and patients with hyperthyroidism may also exhibit low cholesterol.
Carotene	Nonspecific indicator of malabsorption and poor nutritional intake.
Cytokines	Research is exploring prognostic use of cytokine measurements as indicators of inflammatory status.
Electrolytes, blood urea nitrogen (BUN), creatinine, and glucose	Monitor for abnormalities consistent with under- or overhydration status and purging (contraction alkalosis). BUN may also be low in the setting of markedly reduced body cell mass. BUN and creatinine are elevated in renal failure. Hyperglycemia may be nonspecific indicator of inflammatory response.
Complete blood count with differential	Screen for nutritional anemias (iron, B_{12} , and folate), lymphopenia (malnutrition), and thrombocytopenia (vitamin C and folate). Leukocytosis may be observed with inflammatory response.
Total lymphocyte count	Relative lymphopenia (total lymphocyte count $<1200/\mu\text{L}$) is a nonspecific marker for malnutrition.
Helper/suppressor T-cell ratio	Ratio may be reduced in severely undernourished patients. Not specific for nutritional status.
Nitrogen balance	24-h urine can be analyzed for urine urea nitrogen (UUN) to determine nitrogen balance and give indication of degree of catabolism and adequacy of protein replacement. Requires accurate urine collection and normal renal function. Nitrogen balance = $(\text{protein}/6.25) - (\text{UUN} + 4)$. Generally negative in the setting of acute severe inflammatory response.
Urine 3-methylhistidine	Indicator of muscle catabolism and protein sufficiency. Released upon breakdown of myofibrillar protein and excreted without reutilization. Urine measurement requires a meat-free diet for 3 days prior to collection.
Creatinine height index (CHI)	CHI = $(24\text{-h urinary creatinine excretion}/\text{ideal urinary creatinine for gender and height}) \times 100$. Indicator of muscle depletion. Requires accurate urine collection and normal renal function.
Prothrombin time/international normalized ratio (INR)	Nonspecific indicator of vitamin K status. Prolonged in liver failure.
Specific micronutrients	When suspected, a variety of specific micronutrient levels may be measured: thiamine, riboflavin, niacin, folate, pyridoxine, vitamins A, C, D, E, B_{12} , zinc, iron, selenium, carnitine, and homocysteine—indicator of B_{12} , folate, and pyridoxine status.
Skin testing—recall antigens	Delayed hypersensitivity testing. While malnourished patients are often anergic, this is not specific for nutritional status.
Electrocardiogram	Severely malnourished patients with reduced body cell mass may exhibit low voltage and prolonged QT interval. These findings are not specific for malnutrition.
Video fluoroscopy	Helpful to evaluate suspected swallowing disorders.
Endoscopic and x-ray studies of gastrointestinal tract	Useful to evaluate impaired function, motility, and obstruction.
Fat absorption	72-h fecal fat can be used to quantitate degree of malabsorption.
Schilling test	Identify the cause for impaired vitamin B_{12} absorption.
Indirect calorimetry	Metabolic cart can be used to determine resting energy expenditure (REE) for accurate estimation of energy needs. Elevated REE is a sign of systemic inflammatory response.

Source: Reproduced with permission from GL Jensen: *Nutritional Syndromes*. In: D Korenstein (Ed). ACP Smart Medicine [publisher archive]. Philadelphia (PA): American College of Physicians, 2013.

hypersensitivity testing may be measured to demonstrate improvements with nutritional repletion, though it must be appreciated that these are multivariable outcomes for which improved nutritional status is but one variable.

FURTHER READING

- C T et al: ESPEN guidelines on definitions and terminology of clinical nutrition. *Clin Nutr* 36:49, 2017.
- D AS et al: What is Subjective Global Assessment of nutritional status? *J Parenter Enteral Nutr* 11:8, 1987.
- G RS et al: Usefulness of six diagnostic and screening measures for undernutrition in predicting length of hospital stay: A comparative analysis. *J Acad Nutr Diet* 115:927, 2015.
- J GL: Inflammation as the key interface of the medical and nutrition universes: A provocative examination of the future of clinical nutrition and medicine. *J Parenter Enteral Nutr* 30:453, 2006.
- J GL: Malnutrition and inflammation—“Burning down the house”: Inflammation as an adaptive physiologic response versus self-destruction? *J Parenter Enteral Nutr* 39:56, 2015.
- J GL et al: Adult starvation and disease-related malnutrition: A proposal for etiology-based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. *J Parenter Enteral Nutr* 34:156, 2010.
- J GL et al: Adult nutrition assessment tutorial. *J Parenter Enteral Nutr* 36:267, 2012.
- J GL et al: GLIM Criteria for the diagnosis of malnutrition: A consensus report from the global clinical nutrition community. *J Parenter Enteral Nutr* 43:32, 2019.
- K H et al: Global Leadership Initiative on Malnutrition (GLIM): Guidance on validation of the operational criteria for the diagnosis of protein-energy malnutrition in adults. *J Parenter Enteral Nutr* 44:992, 2020.
- W JV et al: Consensus statement: Academy of Nutrition and Dietetics and American Society for Parenteral and Enteral Nutrition. Characteristics recommended for the identification and documentation of adult malnutrition (under-nutrition). *J Parenter Enteral Nutr* 36:275, 2012.

composition, and the details of providing it. To follow these steps properly, physicians require a general understanding of nutritional physiology, nutrient requirements, and the pathophysiology and diagnosis of the nutritional disorders, and familiarity with the indications, advantages, risks, and administration of the different kinds of SNS. Because most physicians are incompletely trained in clinical nutrition, they must collaborate with clinical dietitians and specialized pharmacists in this process.

NUTRITIONAL PHYSIOLOGY (See Chaps. 332-334)

Energy Total daily energy expenditure (TEE) of a healthy sedentary adult is ~36 kcal/kg. Resting energy expenditure (REE), which accounts for ~75% of TEE, may be measured by indirect calorimetry or estimated using a variety of predictive equations that input weight, height, age, sex, and sometimes, disease-related factors. Fever and some forms of critical illness increase REE, whereas prolonged semi-starvation induces an adaptive reduction in REE and voluntary physical activity. Patients' TEE identifies the amount of dietary energy they must consume and metabolize to maintain their existing store of body fat (and protein). The amount of energy a patient *requires* may be less than TEE (in obesity therapy or, temporarily, during periods of energy intolerance) or greater than TEE (during recovery from starvation disease).

Protein and Amino Acids Dietary protein must be consumed throughout life because endogenous protein turnover entails a minimum obligatory rate of amino acid catabolism. Amino acid catabolism increases and decreases in response to changes in protein intake, but it cannot fall below a certain minimum rate that determines an individual's minimum dietary protein requirement. The average daily minimum protein requirement of a healthy adult is 0.65 g/kg; the “safe” or “recommended” intake is 0.80 g/kg. The average protein consumption in wealthy societies is approximately twice the average minimum requirement.

Many diseases (or their treatments) increase the protein requirement, by (1) increasing amino acid loss from the body (as in malabsorption and protein loss via wound exudates, fistulas, or inflammatory diarrhea), removing amino acids from the circulation (renal replacement therapy), or (2) increasing muscle protein catabolism, as occurs as a side effect of high-dose glucocorticoid therapy and especially as part of the metabolic response to systemic inflammation. Highly protein-catabolic patients may excrete 15 g N (nitrogen)/d or more in their urine in the absence of dietary protein provision—this is more than three times faster than during simple fasting. Since 1 g N lost from the body reflects the loss of 6.25 g formed protein, 15 g N loss/d indicates the loss of $15 \times 6.25 = 94$ g protein/d; since the body's metabolically active tissue mass (its body cell mass, 80% of which is skeletal muscle) is ~20% protein, 94 g protein loss/d indicates the loss from the body of ~470 g (1 lb) of muscle mass per day! Sufficiently generous protein provision can reduce this kind of muscle atrophy. The extent to which protein-catabolic illness increases the protein requirement is debated, but the most frequent current recommendation for critically ill patients is 1.5 g protein/kg normal body weight per day—close to the habitual protein intake of healthy people in wealthy societies.

Protein-Energy Interactions Energy deficiency and systemic inflammation increase the dietary protein requirement. Systemic inflammation reduces, but does not prevent, the beneficial effect of increased protein provision during energy deficiency, so long as there is a minimum supply of energy, such as 50% of TEE. Energy provision >50–70% TEE has little further protein-sparing effect in systemic inflammation, and the additional amounts of glucose and fluid volume required to deliver it can have adverse effects.

Permissive Underfeeding and Hypocaloric Nutrition These terms have different meanings, and they should not be conflated or confused. Permissive underfeeding is the deliberate underprovision of all nutrients, including protein, whereas hypocaloric nutrition is energy provision deliberately set less than TEE with a compensatory increase in protein provision.

335

Enteral and Parenteral Nutrition

L. John Hoffer, Bruce R. Bistrian,
David F. Driscoll

There are three kinds of specialized nutritional support (SNS): (1) optimized voluntary nutritional support, which is indicated when a patient's barriers to adequate nutrition can be overcome by special attention to the details of how their food is constituted, prepared, and served and its consumption monitored; (2) forced enteral nutrition (EN), in which a liquid nutrient formula is delivered through a tube placed in the stomach or small intestine; and (3) parenteral nutrition (PN), in which a nutritionally complete mixture of crystalline amino acids, glucose, lipid emulsions, minerals, electrolytes, and micronutrients is infused directly into the bloodstream.

When does a hospitalized patient need SNS? When SNS is indicated, how should it be provided? This chapter summarizes the physiologic principles that guide the correct use of SNS and offers practical information about the diagnosis and management of nutritional disorders in adult hospitalized patients.

The management of in-hospital nutritional disorders follows three steps: (1) screening and diagnosis; (2) determination of the severity and urgency of treating a diagnosed nutritional disorder in its overall clinical context; and (3) selection of the modality of SNS, its

2540 **Micronutrients** Minimum amounts of the nine water-soluble vitamins (the B vitamins and vitamin C), four fat-soluble vitamins (A, D, E, and K), eight minerals (calcium, phosphorus, potassium, sodium, chloride, magnesium, zinc, and iron), essential fatty acids, and several essential trace elements are required to avoid deficiency diseases. Overt deficiencies of potassium, sodium, magnesium, and phosphorus occur so frequently in hospitalized patients that it is standard practice to monitor for and correct them. Certain drugs induce renal potassium, magnesium, or zinc losses that necessitate appropriate increases in their provision. Gastrointestinal losses from nasogastric drainage tubes or intestinal losses from fistulas or diarrhea incur losses of potassium, sodium, calcium, magnesium, and zinc that increase their daily requirement.

Less studied, but common, are subclinical deficiencies of zinc, vitamin C, vitamin D, and possibly other micronutrients. Physicians often assume that consumption of a regular hospital diet protects patients from these deficiencies. This assumption is not warranted when the patient's nutritional status was deficient when they were admitted to hospital and remains so during their hospital stay.

■ MACRONUTRIENT MALNUTRITION SYNDROMES

The decision to embark on SNS must be justified by a well-formulated nutritional diagnosis and clearly defined therapeutic goals. This chapter focuses on the diagnosis, treatment, and prevention of in-hospital starvation-related malnutrition (SRM) and two related conditions: chronic disease-related malnutrition (CDM) and acute disease-related (or injury-induced) malnutrition (ADM). As explained in [Chap. 334](#), SRM results solely from prolonged semi-starvation. CDM is usefully understood as SRM (i.e., simple starvation) that is complicated by moderately severe systemic inflammation. SRM and CDM are anatomically (phenotypically) similar but etiologically and metabolically distinct variations of starvation disease. ADM refers to an injury-induced metabolic condition that creates a high risk of severe body protein deficiency, rather than to an already-existing anatomic starvation disease.

Starvation-Related Malnutrition The pathologic features that define SRM—and distinguish it from the semi-starvation that precedes it—emerge when the body cell mass has been depleted enough to impair specific physiologic functions. Other terms for SRM are “starvation-induced protein-energy malnutrition,” “starvation disease,” and “hunger disease.”

The body normally adapts to starvation by reducing REE and net protein catabolism, partly by means of hormone- and nervous system-regulated changes in cellular metabolism and partly by reducing its muscle mass. These adaptations allow prolonged survival, but survival comes at a cost that includes muscle atrophy (including of the cardiac and respiratory muscles), skin thinning, lethargy, a tendency to hypothermia, and functional disability. The cardinal anatomic diagnostic features of SRM—generalized muscle atrophy and subcutaneous adipose tissue depletion—are easily identified by simple physical examination.

SRM always manifests as weight loss, but weight loss alone may not reveal its full severity. Semistarvation increases the extracellular fluid (ECF) volume (and body weight), sometimes seriously enough to cause edema (“starvation edema”). In adults with initially normal body composition, starvation-induced weight loss tracks the loss of body cell mass (since weight change due to reductions in adipose tissue and increases in ECF volume tend to cancel one another out). A 25% reduction in body weight significantly compromises physiologic function; a 50% reduction places otherwise uncompromised young adults on the cusp of thermodynamic survival; older patients with comorbidities are at even greater risk. People with SRM feel unwell, lack strength, are frail, and are at risk of hypothermia.

The main cause of SRM worldwide is involuntary food deprivation; its causes in hospitalized patients are many. They include inadvertent or physician-ordered food deprivation; psychologic depression or distress; anorexia nervosa; poorly controlled pain or nausea; badly presented unappealing food; communication barriers; physical or sensory disability; dysphagia and other mechanical difficulties ingesting

food; partial obstruction of the esophagus, stomach, or intestinal tract; thrush; intestinal angina; and most commonly, combinations of these causes.

Chronic Disease-Related Malnutrition and Cachexia These terms refer to SRM complicated by chronic systemic inflammation. CDM is prevalent among patients with chronic infection, inflammatory autoimmune disease, chronic severe hepatic, renal, cardiac, and pulmonary disease, and neoplastic diseases that induce a systemic inflammatory response or cause tissue injury. CDM causes and is worsened by anorexia—a strong disinclination to eat even when there is no physical barrier to it—and is characterized by an increased rate of muscle protein catabolism, muscle atrophy, weakness, fatigue, and a subverted adaption to starvation, all of which contribute to a vicious cycle of worsening disease. Fortunately, the nutritional deficit on the input side of this equation (anorexia-driven inadequate food consumption) is often a stronger driver of the patient's CDM than increased nutrient loss on the output side (increased amino acid catabolism and, sometimes, increased energy expenditure). This makes CDM amenable to a well-organized nutritional intervention while effective treatment of the primary disease is implemented. The challenge becomes more daunting when there is no effective therapy for the primary disease.

Cachexia is an older term that refers to a disease-induced metabolic syndrome characterized by moderate systemic inflammation, unrelenting and severe generalized muscle atrophy, and the symptoms associated with them; it is, therefore, approximately synonymous with CDM. Anyone with cachexia has CDM, but in the view of some clinicians, CDM that is milder and less sustained does not qualify for the term.

Acute Disease-Related Malnutrition Other terms for ADM are “injury-induced malnutrition” and “protein-catabolic critical illness.” The metabolic-inflammatory response to severe tissue injury and sepsis mobilizes muscle amino acids and leads to rapid and severe generalized muscle atrophy and variable increases in REE under conditions in which voluntary food intake is almost always impossible. SRM or CDM may or may not be present at the onset of their critical illness, but muscle atrophy will rapidly develop or worsen unless the inciting medical or surgical disease is rapidly and effectively treated and SNS provided. The rate of loss of body cell mass in ADM can be three to five times greater than in simple starvation. Death from simple starvation of non-obese adults occurs within ~8 weeks; death due to untreated starvation of patients with sustained ADM will occur correspondingly sooner.

■ NUTRITIONAL DIAGNOSIS

The cardinal anatomic features of starvation disease (SRM or CDM) are generalized muscle atrophy and diminished body fat. A routine physical examination will reveal these features, but what should be an easy diagnosis is often overlooked. This section explains the details and pitfalls of diagnosing SRM and CDM.

Muscle Mass Generalized muscle atrophy is easy to identify, and its severity is determinable almost at a glance. Serum creatinine adjusted for renal function or urinary excretion, adjusted for height and sex, may also confirm severe muscle atrophy. A problem with diagnosing SRM and CDM is that muscle atrophy has many causes. They include (1) old age-related muscle atrophy (sarcopenia); (2) disuse muscle atrophy; (3) high-dose glucocorticoid therapy and certain endocrine diseases (uncontrolled diabetes mellitus, adrenocortical insufficiency, hyperthyroidism, androgen deficiency, hypopituitarism); and (4) primary muscle or neuromuscular diseases. The guiding clinical principle is that SRM and CDM are extremely common causes of and contributors to muscle atrophy. Whenever generalized muscle atrophy is observed, its potential causes should be evaluated and the treatable ones addressed. Old age is irreversible, but adequate protein and energy provision to starving patients, combined with physical rehabilitation for immobile patients, can be lifesaving.

Generalized muscle atrophy of any cause is especially dangerous in ADM, because patients suffering from ADM and muscle atrophy are closer to the cliff edge of lethal depletion of their body cell mass. In addition, a diminished muscle mass is less able to release adequate

amounts of amino acids into the circulation for protein synthesis at sites of tissue injury and healing and to the central protein pool to regulate the immunoinflammatory response.

Subcutaneous Adipose Tissue Severe adipose tissue depletion indicates starvation disease, but it need not be present to make the diagnosis. The current obesity epidemic has created a population of obese patients with SRM or CDM in whom muscle atrophy has outpaced fat loss. A conscientious physical examination easily identifies these patients' atrophic muscles despite their residual subcutaneous fat.

ECF Volume The ECF volume normally represents ~20% of body weight. SRM moderately increases ECF volume. Patients with CDM have additional edema-promoting conditions, especially hypoalbuminemia. Unless the effect of ECF is accounted for, its increased volume may conceal the severity of muscle atrophy in patients with SRM and CDM.

Body Mass Index Body mass index (BMI) is defined as body weight (kg) divided by the square of height (m^2). BMI normally ranges from 20 to 25 kg/m^2 ; values <19–20 usually indicate reduced muscle and fat mass. BMI <15 indicates severe starvation disease; BMI <13 is usually thermodynamically incompatible with life, especially in older patients with comorbidities. Some guidelines and clinical trial enrollment criteria define “malnutrition”—in this context, a synonym for starvation disease—as a BMI <16 or 17. Using these criteria alone can lead to serious error. A BMI <17 certainly indicates starvation disease—the body architecture associated with such a BMI can only be created by jettisoning a large fraction of the body cell mass and adipose tissue store. But a BMI >17 does not rule out starvation disease. Many patients with starvation disease have a normal or above-normal BMI despite their muscle atrophy because of residual obesity or an expanded ECF volume.

Visual BMI After some practice and verification, clinicians can accurately predict the BMI of nonobese, nonedematous patients by attentively examining their muscle groups. Once acquired, this skill enables them to interpolate the severity of starvation disease in obese or edematous patients—in whom measured BMI is unreliable—by evaluating their muscles while intuitively discounting their subcutaneous fat and edema. Visual BMI may also be used to estimate a patient's normalized dry body weight (i.e., weight adjusted for obesity, edema, or ascites). For example, the normalized dry body weight of a 1.75-m adult with a visual BMI of 17 is $1.75^2 \times 17 = 52$ kg. Since protein and energy targets are based on the patient's body normalized weight, this calculation is useful when body weight is unreliable or difficult to measure.

Laboratory and Technical Assessment Clinical laboratory measurements have three main purposes in the evaluation and management of starvation disease.

MUSCLE MASS Bedside ultrasound is a potentially valuable technique for quantifying muscle mass at specific body sites, but it need not, nor should it, replace the comprehensive evaluation provided by the eyes, hands, and discerning mind of a bedside examiner.

SYSTEMIC INFLAMMATION The absence or presence of systemic inflammation distinguishes SRM from CDM. The most useful laboratory indicators of systemic inflammation are a reduced serum albumin concentration and increased serum C-reactive protein concentration. Systemic inflammation increases the permeability of capillary walls to large molecules; the resulting osmotic shift increases the ECF volume. Intravascular albumin redistributes into this larger volume, decreasing the serum albumin concentration (increased albumin catabolism also contributes). Dietary protein deficiency and muscle atrophy combine to perpetuate inflammation-induced hypoalbuminemia, because the amino acids used for hepatic albumin synthesis are derived from the diet and endogenous muscle protein.

Hypoalbuminemia and reduced serum prealbumin concentrations are often claimed to diagnose “malnutrition.” This is incorrect. Serum albumin and prealbumin are negative acute-phase reactants

that indicate systemic inflammation. Systemic inflammation induces anorexia and increases muscle catabolism, increasing the risk of CDM, but the disease itself may or not exist at the time and may never develop. The serum concentrations of acute-phase reactants will not improve while systemic inflammation persists, even with prolonged optimal nutritional therapy.

PROTEIN CATABOLIC INTENSITY The defining feature of protein-catabolic disease (which occurs in a moderate form in CDM and severely in ADM) is increased net muscle amino acid catabolism. Conditions that substantially increase body protein loss can be identified by measuring body N loss. Most N leaves the body in the urine (almost all of it in urea, ammonium, and creatinine). Total N is not usually measured in hospital laboratories, but the analysis of urinary urea N (which normally accounts for ~85% of urinary N) is routinely available. A recent, validated formula estimates daily total N loss as follows: N loss (g) = g N in urinary urea/0.85 + 2.

Net muscle protein catabolism follows approximately first-order kinetics, such that the rate of N loss from muscle is proportional to the existing amount of N available to be lost. Muscle-atrophic, protein-catabolic patients lose less body N per day than equivalently catabolic patients whose muscle mass is normal, but they are nevertheless at greater risk of succumbing to their critical illness. The interpretation of a patient's rate of N loss should include an evaluation of their existing muscle mass.

Instrumental Nutritional Assessment Many nutritional assessment instruments claim to identify “malnutrition” by enumerating and summing a list of risk factors, laboratory results, and physical findings. These tools are often hindered by ambiguity about the definition of malnutrition and by failure to distinguish between screening and diagnosis. Diagnosis is the process of identifying a known pathologic entity—SRM or CDM, for example—by considering the patient's medical history, pertinent findings on physical examination, and laboratory or imaging reports. Diagnosis also involves an estimation of the probability that the diagnosis is correct and a judgment of its severity. By contrast, screening is the application of a simple test that identifies people at sufficiently high risk of a certain disease to warrant definitive procedures to establish the diagnosis or rule it out or that identifies people at sufficiently high risk of developing the disease to warrant specific preventive interventions. Screening tools and risk predictors are useful, but it is a mistake to confuse them with clinical diagnosis.

SPECIALIZED NUTRITIONAL SUPPORT

Optimized Voluntary Nutritional Support When feasible, this is the approach of choice because it engages and empowers the patient, encourages mobilization and reconditioning, is consistent with the objectives of patient-centered medicine, and is risk-free. Its disadvantage is that it is time-consuming and labor-intensive, and it demands interest in and attention to the specific needs of individual patients.

Enteral Nutrition This is nutrition provided through a feeding tube placed through the nose into the stomach or beyond it into the duodenum, by insertion of a tube through the abdominal wall into the stomach or beyond it into the jejunum, or by an open surgical approach to access the stomach or small intestine. EN is the treatment of choice when optimized voluntary nutritional support is impossible or has failed. It is relatively simple, safe, and inexpensive and maintains the digestive, absorptive, and immunologic barrier functions of the gastrointestinal tract. EN is appropriate when optimized voluntary nutrition is not feasible or has failed and the patient's gastrointestinal tract is functioning and can be accessed.

EN Products The most common forms of EN used are commercially manufactured formulas with defined compositions.

STANDARD POLYMERIC FORMULAS These are the most widely used sources of EN. They are available in a wide variety of formats that generally meet the nutritional requirements of a normal, healthy person. Carbohydrates provide most of the energy. The proteins in them (casein, whey, or soy) are intact and require normal pancreatic enzyme

function for digestion and absorption. They are isotonic or nearly so and provide 1000–2000 kcal and 50–70 g protein/L.

POLYMERIC FORMULAS WITH FIBER The addition of dietary fiber to formulas sometimes improves bowel function and feeding tolerance. Fermentable (soluble) fibers such as pectin and guar are metabolized by colonic bacteria, yielding short-chain fatty acids that fuel colonocytes. Nonfermentable (insoluble) fibers increase fecal bulk, improve peristalsis, and may improve diarrhea.

ELEMENTAL AND SEMI ELEMENTAL FORMULAS The macronutrients in these formulas are partially or completely hydrolyzed. They are primarily designed for patients with known malabsorption and malabsorption, but they are sometimes used empirically for patients who have had prolonged bowel rest or are critically ill without strong evidence of their superiority, or when a patient is intolerant of a standard polymeric formula.

IMMUNE ENHANCING FORMULAS In addition to providing macronutrients and conventional amounts of micronutrients, these EN products contain large amounts of certain nutrients designed to favorably modulate the immune response: arginine and n-3 fatty acids especially, but also various combinations of glutamine, nucleotides, and antioxidants.

PROTEIN ENRICHED FORMULAS Most EN formulas provide calories and protein in a ratio appropriate for a healthy person, whereas protein-enriched formulas provide ~90 g protein and 1000 kcal/L. Originally marketed to meet the increased protein requirement of weight-reducing obese patients, these products are increasingly used to provide protein-catabolic patients with a more generous amount of protein without energy overfeeding. EN can be further protein-enriched by adding flushes of water-soluble powdered protein supplements.

OTHER FORMULAS Various disease-specific EN products are available for patients with diabetes and hepatic, renal, or pulmonary disease. Their use can improve some metabolic endpoints, but there is no definitive evidence that they improve clinical outcomes.

Parenteral Nutrition PN delivers a complete nutritional regimen directly into the bloodstream in the form of crystalline amino acids, glucose, triglyceride emulsions, minerals (calcium, phosphate, magnesium, and zinc), electrolytes, and micronutrients. Because of its high osmolarity (>1200 mOsm/L) and often large volume, PN is infused into a central vein in adults. Ready-to-use PN admixtures typically containing 4–7% hydrated amino acids and 20–25% glucose (with or without electrolytes) are available in two-chamber (amino acids and glucose) or three-chamber (amino acids, glucose, and lipid) bags that are intermixed with vitamins, trace minerals, and additional electrolytes then added just prior to infusion. Although convenient and cost-effective, these products have fixed nutrient compositions and are dosed according to the volume required to meet a patient's energy requirement but not necessarily their protein requirement. In some situations—especially ADM—a more sophisticated approach is justified that uses a computer-controlled sterile compounding to create combinations of amino acids and glucose that meet the precise protein and energy requirements of individual patients.

Amino Acids PN amino acid admixtures vary, but all of them provide appropriate amounts of the essential amino acids and nonessential amino acid N. The hydrated state of the mixed free amino acids in PN solutions reduces their energy density from 4.0 (in formed protein) to 3.3 kcal/g, and it reduces the amount of protein substrate they provide by 17%. For example, 100 g of free mixed amino acids provides 83 g protein substrate and 330 kcal.

Carbohydrate and Lipids The glucose in PN is dextrose monohydrate; its hydrated state reduces its energy density from 4.0 (in formed carbohydrate) to 3.4 kcal/g. Lipid emulsions provide energy (~10 kcal/g) and the essential n-6 and n-3 fatty acids. Traditional lipid emulsions are based solely on soybean oil, but they are giving way to mixed emulsions that include medium-chain triglycerides, n-9 monounsaturated fatty acids, and n-3 fatty acids. Emulsions of pure soybean

oil, a mixture of 80% olive oil and 20% soybean oil, and a mixture of 30% soybean oil, 30% medium-chain triglycerides, 25% olive oil, and 15% fish oil are available in the United States. (A 10% fish oil emulsion is approved for intestinal failure-associated liver disease in neonates and infants.) Fish oil (either as a component of a mixed emulsion or administered separately) may reduce the risk of infections and length of stay in critically ill patients. The complex lipid emulsions are more highly enriched in n-3 fatty acids and/or contain fewer n-6 polyunsaturated fatty acids than soybean lipid, which is more prone to lipid peroxidation and could promote the formation of the proinflammatory n-6 derivatives. Standard lipid infusion rates should not exceed 8 g/h, equivalent to 175 g (1925 kcal)/d in a 70-kg patient; pure fish oil emulsions must be infused at lower rates.

Minerals, Micronutrients, and Trace Elements The default concentrations of electrolytes, minerals, and micronutrients in PN solutions are designed to meet the requirements of a healthy adult. These starting doses must be adjusted to meet the frequently abnormal and often-changing requirements of individual patients. Being unstable, multivitamin mixtures are injected into PN bags just prior to their delivery to the medical unit. Parenteral water-soluble vitamin requirements are greater than standard oral requirements, because hospitalized patients often have vitamin deficiencies or increased requirements and because intravenous administration of vitamins increases their loss in the urine. Ascorbic acid degrades spontaneously in PN solutions, even when light-protected. The amount of vitamin D in currently available intravenous vitamin products is inadequate.

APPROACH TO THE PATIENT

Indications, Selection, and Provision of Specialized Nutritional Support

Most hospitalized patients do not require SNS because they can eat and will improve with appropriate management of their primary disease. Others have a terminal disease whose downward course will not be slowed by SNS. Patients who cannot eat enough hospital food and who have or are at high risk for SRM or CDM are candidates for optimized voluntary nutrition support. When this most desirable approach is inappropriate or impractical or has been properly tried and failed, invasive SNS must be considered. The decision to provide or withhold EN or PN is based on a synthesis of four factors: (1) the determination that nutrient ingestion will likely continue to be inadequate for many days; (2) the patient has important muscle atrophy (of any cause) or fat depletion; (3) the patient's nutrient requirements are increased (as from inflammatory diarrhea, enterocutaneous fistulas or exudates, or a pronounced inflammatory protein-catabolic state); and (4) the reasoned judgment that SNS has a reasonable prospect of improving the patient's clinical outcome or quality of life.

EN THERAPY

EN is indicated when the patient is unable to eat enough food and is unlikely to do so for a long time, their gastrointestinal tract is functional and accessible, and optimized voluntary nutrition is impossible or cannot meet their nutritional requirements. EN is commonly used for patients with impaired consciousness, severe dysphagia, or severe upper gastrointestinal tract dysfunction or obstruction or who need mechanical ventilation. Equally commonly, situations arise in which a patient's voluntary food intake is seriously curtailed by anorexia, unappealing food, nausea, vomiting, pain, distress, delirium, depression, chewing difficulties, mild dysphagia, physical and sensory disability (including dysgeusia), or undiagnosed thrush. In these complicated and difficult situations, the clinical diagnosis of SRM or CDM should tip the decision from optimized voluntary nutrition toward EN or PN.

EN is contraindicated in patients with intestinal ischemia, mechanical obstruction, peritonitis, and gastrointestinal hemorrhage. High-dose pressor therapy is another relative contraindication,

due to the rare but lethal risk of intestinal ischemic injury. Severe coagulopathy, esophageal varices, absent gag reflex, hypotension, paralytic ileus, pancreatitis, diarrhea, and nausea and vomiting are not absolute contraindications, but they increase the risk of complications and make it less likely that EN will succeed in achieving its nutritional goal.

Initiation, Progression, and Monitoring Nasogastric tube feeding may proceed when the patient's gastrointestinal function is adequate with respect to gastric contractility (e.g., nasogastric tube output <1200 mL/d), intestinal contractility (absence of a known or suspected intraabdominal pathologic process and presence of a nondistended abdomen with detectable bowel sounds, although the *absence* of bowel sounds is not, in itself, a contraindication), and adequate colonic function (passage of stools and flatus). After consent has been obtained and the appropriate feeding tube (usually a nasogastric tube for short-term feeding) has been placed and its position verified, the head of the patient's bed is raised to at least 30° and kept raised to reduce the risk of regurgitation. Clinical dietitians ordinarily order the formula and adjust its rate of provision. When a standard polymeric formula is infused, it normally commences at 50 mL/h and is advanced by 25 mL/h every 4–8 h until the goal rate is attained. Elemental formulas are commenced at a slower rate and progress more slowly. Intraoperative bolus feeding is an option (200–400 mL feeding solution infused over 15–60 min at regular intervals with verification of residual gastric contents every 4 h).

Complications and Their Management The most common complications of EN are aspiration of regurgitated or vomited formula, diarrhea, fluid volume and electrolyte derangements, hyperglycemia, nausea, abdominal pain, constipation, and failure to achieve the nutritional goal.

Aspiration Patients with delayed gastric emptying, impaired gag reflex, and ineffective cough are at high risk of aspiration pneumonia. Ventilator-associated pneumonia is mostly caused by aspiration of microbial pathogens in the mouth and throat past the cuffs of endotracheal or tracheostomy tubes, but tracheal suctioning induces coughing and gastric regurgitation. Measures to prevent ventilator-associated pneumonia include elevation of the head of the bed, mouth hygiene and gastrointestinal decontamination, nurse-directed algorithms for formula advancement, and sometimes, postpyloric feeding. EN does not have to be suspended for gastric residual volumes <300–400 mL in the absence of other signs of gastrointestinal intolerance (nausea, vomiting, severe abdominal pain, abdominal distention). Continuous EN is often tolerated better than bolus feeding, and it is the only option during jejunal feeding.

Diarrhea Diarrhea commonly occurs when the patient's bowel function is compromised by disease or drugs (most often, broad-spectrum antibiotics). Once infectious and inflammatory causes have been ruled out, EN-associated diarrhea may be controlled by using a fiber-containing formula or adding an antidiarrheal agent to it. H₂ blockers or proton pump inhibitors may help reduce the net volume of fluid presented to the colon. Since luminal nutrients have trophic effects on the intestinal mucosa, it is often appropriate to persist with tube feeding despite moderate, tolerable diarrhea, even if it necessitates supplemental parenteral fluid support. Except for patients with markedly impaired small-intestinal absorptive function, there are no well-established indications for elemental formulas, but they may be used empirically when diarrhea persists despite the use of fiber-enriched formulas and antidiarrheal agents.

Gastrointestinal Intolerance Abnormally high gastric residual volumes, abdominal distention, pain, and nausea are distressing for patients, increase the nursing workload, and delay the progression of EN. These problems can be avoided or minimized by ensuring normal fluid and electrolyte balance, by preventing severe

hyperglycemia, and, when a patient experiences nausea, vomiting, or abdominal distension, by the judicious use of antiemetic and prokinetic drugs (and sometimes proton pump inhibitors) on a regular—rather than as-needed—basis. Patients with gastroparesis require postpyloric feeding.

Fluid Volume, Electrolyte, and Blood Glucose Abnormalities EN's essential purpose is to provide macronutrients at an appropriate rate. EN also provides standard amounts of fluid, electrolytes, minerals, and micronutrients. They are not designed to manage abnormal fluid volume, electrolyte, and mineral requirements, which vary considerably among different patients and can change rapidly. Blood glucose concentrations should be monitored regularly, and additional measures—including intravenous fluid, electrolyte, and insulin therapy—should be taken to maintain homeostasis.

Failure to Reach the Nutritional Goal EN is frequently delayed or interrupted by diagnostic tests and procedures (including dialysis), physical or occupational therapy, a clogged or pulled out tube, and intolerance to EN. The result can be a long delay in the progression of EN and ultimate failure to meet the patient's nutrient requirements.

EN in the Intensive Care Unit Most critically ill patients cannot eat anything—they depend entirely on SNS. EN serves two purposes in this setting. The first is to meet the patient's macronutrient requirements, especially their often dramatically increased protein requirement. The second purpose is to infuse nutrients into the intestines at a rate that sustains normal intestinal barrier and immunologic functions in the face of a systemic inflammatory response that threatens intestinal integrity and immune function. Current guidelines recommend starting EN soon after a critically ill patient has been fluid resuscitated and stabilized. Once EN is underway, the rate of delivery is increased as tolerated until the patient's nutritional goal is achieved. EN often falls far short of the protein provision target, even after a week or longer in the intensive care unit. Newer, high-protein EN products and the addition of powdered protein supplements can correct this protein shortfall.

PN THERAPY

PN is more resource-intensive, is potentially riskier, and requires more expertise than EN. It is used when invasive SNS is indicated and EN is impossible, inappropriate, or insufficient to meet the patient's nutritional needs. The risks of PN are those of inserting and maintaining a central venous catheter (traumatic injury from the insertion, serious infection, and venous thrombosis); allergy to some of its components; glucose, electrolyte, magnesium, phosphate, and acid-base balance abnormalities; and the adverse effects of the large intravenous fluid volumes. PN that is prolonged for many weeks—especially when it delivers excess energy—may cause or contribute to hepatic dysfunction.

Initiation, Progression, Monitoring, and Discontinuation When indicated, PN should begin as soon as possible after the patient has been hemodynamically resuscitated, glucose, electrolyte, and acid-base homeostasis has been established, and they can tolerate the fluid volume required to deliver it. The high osmolarity of adult PN solutions and need for strict sterility require their infusion through a dedicated port in a central venous catheter. Jugular or femoral vein catheters should not be used because of the difficulty maintaining a dry, sterile dressing over the insertion site. The initial dose of glucose should not exceed 200 g/d to avoid hyperglycemia (and—in susceptible patients with adapted SRM—the refeeding syndrome). The full dose of amino acids can be administered from the very first day—an option that is, unfortunately, unavailable when premixed PN solutions are used.

Most non-critically ill patients (e.g., dry body weight 70 kg) do not require >500 g glucose (1700 kcal/d), and many, if not most, patients with ADM do not require >350 g (1200 kcal/d) during the intense phase of their disease. A glucose infusion rate of ~200 g/d

is physiologic and commonly does not have to be exceeded. When it eventually becomes appropriate to set the energy goal equal to TEE, it may be achieved by infusing a lipid emulsion. Even lower glucose infusion rates (e.g., 100–200 g/d) are safe during deliberate hypocaloric nutrition and may prevent or minimize hyperglycemia in insulin-resistant patients.

We recommend hypocaloric nutrition (high in protein but limited in glucose, lipid, and fluid volume) for the first 2 weeks of SNS in fat-sufficient or obese patients with ADM. Energy provision can increase, if indicated, after the catabolic storm abates. Lipids are commonly introduced after the first week of PN and can be used to make up energy shortfalls. Serum triglyceride concentrations are measured before commencing lipid infusions to detect preexisting hypertriglyceridemia (>400 mg/dL)—a relative contraindication. Lipids may be infused daily or two to three times weekly. Lipid infusions are not necessary to prevent essential fatty acid deficiency during hypocaloric nutrition of obese patients, because the mobilization of body fat during energy deficiency provides the body with endogenous essential fatty acids.

Capillary blood glucose concentrations are monitored several times daily, and subcutaneous regular insulin is added to the PN admixture as required to maintain average serum glucose concentrations <140 mg/dL and >80 mg/dL. (Upper and lower limits of 180 and 100 mg/dL appear to be appropriate for critically ill patients with diabetes mellitus.) The dose of regular insulin required on a given day can be added to the following day's PN solution. The insulin dose increases roughly proportionately to the glucose dose. Certain benchmarks are useful. Basal endogenous insulin secretion is ~30 units/d in normal people. When insulin is required for nondiabetic, noncatabolic patients, 10 units of regular insulin roughly cover 100 g infused glucose. Patients with non-insulin-dependent diabetes require ~20 units/100 g glucose. Noncatabolic patients with insulin-dependent diabetes usually require approximately twice the at-home insulin dose, because parenteral glucose stimulates insulin release more potently than oral carbohydrate and because some insulin adheres to the infusion bag.

Biochemical Monitoring Serum urea, creatinine, electrolytes, glucose, magnesium, phosphate, calcium, and albumin concentrations are measured prior to starting PN and followed daily for the first few days, then twice weekly or as required. Serum triglycerides and liver function tests (and often ferritin) are measured at baseline and after PN is underway to confirm that the lipid infusions are well tolerated. N balance, calculated from 24-h urinary urea N excretion, is useful at the outset for evaluating the severity of protein catabolism in patients with CDM or ADM, to identify patients who require more generous amino acid provision, and during PN to determine whether the patient's N balance is improving with therapy.

Discontinuation PN is tapered and discontinued when the patient can be adequately nourished by the enteral route. The dose of PN is gradually reduced as food intake increases. Once a patient is tolerating one-half to two-thirds of their food requirement by the enteral route and there is no mechanical or other barrier to further increases in intake, PN should be terminated. The transition to oral nutrition can be slow for patients with CDM. In this situation, optimized voluntary nutrition, although labor-intensive, is much preferred to replacing PN with invasive EN because it is safe, effective, fosters well-being, and prepares patients for discharge home. The temptation to discontinue PN to stimulate a patient to eat more food should, in general, be resisted. PN does not create anorexia, nor does discontinuing it stimulate appetite. Too-early discontinuation of PN may delay a patient's progression to full voluntary food consumption by inducing anxiety and recreating starvation conditions. A patient is most successfully weaned from PN by optimizing their voluntary nutrition (including food from home), providing emotional support, encouraging physical activity, and being patient. Some patients, stuck on the cusp of adequate oral nutrition, will

benefit from discharge to the security and pleasure of home life and homemade food; these patients are identified by observing, asking, and listening.

Drawbacks, Side Effects, and Complications Patients receiving PN are at greater risk of bloodstream infections than other patients with central venous catheters. Rigorously aseptic insertion technique, meticulous dressing care, one port dedicated solely to PN, and careful glycemic control reduce this risk.

Hyperglycemia The most frequent metabolic complication of PN is hyperglycemia in patients with insulin resistance due to non-insulin-dependent diabetes mellitus, high-dose glucocorticoid therapy, or severe systemic inflammation; the problem is exacerbated by excessively high rates of glucose provision. Glucose concentrations are most easily kept at <140 mg/dL with the least risk of hypoglycemia by infusing hypocaloric amounts of glucose and, when necessary, meeting the patient's energy requirement with intravenous lipid. In ADM, the benefits of using the lowest possible insulin dose—minimal hyperinsulinemia and a reduced risk of hypoglycemia—almost always outweigh the doubtful goal of rapidly matching energy provision to the energy expenditure rate of patients whose existing fat store is normal.

Hypoglycemia Reactive hypoglycemia is uncommon but may occur when high-glucose, non-insulin-containing PN is abruptly discontinued. It is prevented by slowing the PN infusion rate to 50 mL/h for 1 or 2 h prior to discontinuing it (or replacing it with 10% glucose) or, when the oral route is available, providing a snack. More often, hypoglycemia occurs when the intensity of the patient's metabolic stress (or their glucocorticoid dose) decreases without an appropriate downward adjustment of the insulin dose. This problem is avoided by frequent capillary glucose determinations and careful attention to medication doses and the patient's general condition.

Artefactual Hyperglycemia and Hyperkalemia Blood samples must be meticulously collected from a dual-port central venous catheter. Intermixing of the sample with even a tiny volume of PN solution will falsely indicate hyperglycemia and hyperkalemia and may trigger a treatment error. The sampling error is identified when the patient's apparent serum glucose (and potassium) concentrations abruptly increase without reason and the apparently very high glucose concentration is out of keeping with concurrent capillary glucose readings.

Volume Overload Hypertonic intravenous glucose triggers a more intense insulin response than oral glucose that can increase urinary sodium and water retention. In this setting, net fluid retention is likely when total fluid provision exceeds 2 L/d in patients not experiencing large gastrointestinal losses. The problem of volume overload can be minimized by using a compounding to prepare PN solutions, infusing glucose at a rate that minimizes the need for exogenous insulin therapy, and avoiding energy overfeeding.

Hypertriglyceridemia This complication occurs when the rate of lipid infusion exceeds plasma triglyceride clearance capacity. Sepsis, renal failure, diabetes mellitus, high-dose glucocorticoid therapy, and multiple-organ failure reduce triglyceride clearance. An impaired immune response, increased risk of acute pancreatitis, and altered pulmonary hemodynamics are potential, but not well documented, complications of PN-induced severe hypertriglyceridemia. Lipid infusion rates should not usually exceed ~50 g (500 kcal)/d in ADM.

Liver Disease Mild elevations of serum liver enzyme concentrations can occur within 2–4 weeks of initiating PN, but in most cases, they return to normal even when PN is continued. Clinically important hepatic dysfunction, although common in children, is uncommon in adults when energy overfeeding and resultant fatty liver are avoided. Intrahepatic cholestasis occasionally occurs after many weeks of continuous PN and is most often multifactorial in

origin. Cyclic PN—in which PN is infused for only 12 h of the day—may prevent or reduce the severity of this complication.

PN in the Intensive Care Unit Current guidelines recommend starting EN soon after a critically ill patient has been resuscitated, stabilized, and enteral access established to an adequately functioning gastrointestinal tract. EN is then advanced over the following days. If the energy goal has not been achieved after 7–10 days, PN is recommended, especially if the patient's protein-catabolic state has not yet abated. Soy-based lipid emulsions should be avoided during the first week of PN during critical illness; alternative lipid emulsions may prove to be safe and beneficial.

SPECIAL CLINICAL SITUATIONS

Critical Illness–Nutrition Paradox High-quality evidence now confirms what has long been indicated by the biologic evidence, physiologic reasoning, formal observational studies, and objective clinical observation, namely, that personalized nutritional interventions improve the clinical outcomes of starving, non–critically ill patients. The case for SNS would appear to be even stronger in ADM—with its rapid, severe muscle atrophy and maintained or increased energy expenditure under conditions in which patients are almost always unable to eat voluntarily—but well-designed clinical trials of nutritional interventions in critical illness have repeatedly failed to demonstrate that currently prescribed SNS regimens improve the clinical outcomes of critically ill patients. The evidence does indicate that, unlike in noncritical illness, energy provision that is set at or near the rate of energy expenditure in fat-sufficient, insulin-resistant critically ill patients does not improve their clinical outcomes and may be deleterious to some of these patients. The inability of currently prescribed SNS to improve outcomes in critical illness has several possible explanations: (1) severe prolonged starvation is so harmful to all people, whether critically ill or not, that ethical considerations preclude using deliberate starvation as a treatment arm in a clinical trial; (2) critical illness is enormously heterogeneous, and not every critically ill patient is or remains severely protein-catabolic for long; (3) owing to more generous admission criteria and thanks to the high quality of modern intensive care, many patients admitted to intensive care units improve and are discharged within a handful of days, whereas others are so mortally ill that their clinical outcome is virtually predetermined, and proof-of-concept clinical trials that enroll and report the outcomes of such patients could fail to demonstrate a benefit from SNS; and (4) in current practice, the EN-based SNS regimens that are prescribed for most critically ill patients commonly fail to deliver more than one-half the currently recommended amount of protein. The low protein-to-energy ratio of most standard EN and PN products makes it difficult to provide critically ill patients with a sufficiently generous amount of protein or amino acids while avoiding energy overfeeding. (The problem can be exacerbated by use of the sedative drug propofol, which is infused in a solution of 10% lipid that commonly delivers ~500 kcal/d.) For these reasons, together with other experts, we continue to recommend EN and PN for critically ill patients with ADM, with the additional advice to avoid energy overfeeding during the initial weeks (or as long as systemic inflammation remains severe) by deliberately erring on the side of hypocaloric nutrition while simultaneously providing suitably generous protein or amino acids, as guided by physiologic reasoning and a personalized evaluation of the anatomic and etiologic-metabolic condition of each patient.

Iron and PN Iron deficiency is common in hospitalized patients; its usual causes are preexisting deficiency, inadequate in-hospital dietary provision, macro- or microscopic gastrointestinal blood loss, and repeated blood sampling. The diagnosis is often missed because the anemia of systemic inflammation is much more common, and it increases serum concentrations of ferritin, a positive acute-phase reactant. Iron is not routinely added to PN mixtures. Iron dextran is incompatible with lipid emulsions, and although

it appears to be chemically compatible with aqueous solutions of amino acids and glucose, there is realistic concern that interactions between iron molecules and certain vitamins and amino acids in PN solutions could catalyze the formation of free radicals that degrade vitamins and exert subtle adverse systemic effects. In principle, all micronutrient deficiency states, including iron deficiency, should be prevented and corrected. In-hospital iron deficiency causes and prevents recovery from anemia, and subclinical iron deficiency could contribute to cognitive and immune dysfunction. Serum ferritin concentrations should be determined when PN commences and remeasured at approximately 8-week intervals. Iron deficiency is strongly suggested by an intermediate serum ferritin concentration in the setting of systemic inflammation and by decreasing mean red cell volumes (even within the low-normal range). Intravenous iron should be administered according to standard guidelines. A termination order should be written to prevent inadvertent iron overdosing. Parenteral iron therapy should be avoided in ADM because a substantial rise in the serum iron concentration could release free iron and increase susceptibility to gram-negative (and possibly other microbial) infections, as well as catalyze the formation of free radicals that increase the intensity of the catabolic response to major tissue injury.

Zinc One liter of secretory diarrhea contains ~12 mg of zinc. Patients with intestinal fistulas or high-volume chronic diarrhea require this amount of zinc in addition to their daily requirement of 15 mg to avoid zinc deficiency. Zinc may be provided parenterally or enterally. Because of its low bioavailability, 12 mg of parenteral zinc is equivalent to 30 mg of oral zinc.

Old Age In addition to their other frailties, elderly people commonly suffer from age-related muscle atrophy (sarcopenia) compounded by disuse muscle atrophy. These factors place them at high risk of the consequences of starvation disease and make them candidates for early SNS.

Inactivity Physical activity and adequate nutrition are closely interdependent. Reduced physical activity reduces appetite, and physical rehabilitation and its associated emotional benefits restore optimism and appetite. Full nutrient provision will maintain or normalize many physiologic functions in bedridden patients, but they will not increase muscle mass.

Renal Failure Protein provision should not be reduced in patients with renal failure unless renal replacement therapy is unavailable. Renal replacement therapy removes large amounts of amino acids, vitamins, and trace elements from the circulation, so protein and micronutrient provision should be increased to compensate for these losses.

Liver Failure Patients with severe hepatic disease are relatively intolerant to starvation and commonly have CDM when admitted to hospital, so they are prime candidates for SNS. Their SNS should be generous both in energy and protein, despite an increased risk of hepatic encephalopathy. The risk of encephalopathy can be reduced by meticulous attention to fluid balance, acid-base balance, and electrolyte status and by spreading protein provision over the day to accommodate the liver's reduced capacity to clear amino acid-derived ammonia.

Perioperative SNS Patients with SRM or CDM awaiting elective major surgery benefit from 7–10 days of preoperative SNS. When feasible and properly implemented, optimized voluntary nutrition is greatly to be preferred, but when a patient has been admitted to hospital in a semi-urgent condition, EN or PN will meet the patient's nutritional goal more quickly. Preoperative SNS improves immunity and reduces postoperative complications, but it will not increase serum albumin concentrations, and it should not be provided for >7–10 days with that goal in mind. More prolonged preoperative EN or PN may confer slight additional nutritional benefits, but they are counterbalanced by their risks and the consequences

of prolonged hospitalization and delayed surgery. Surgery should not be delayed for starving patients whose muscle mass is normal or only mildly depleted and who are not experiencing systemic inflammation since they tolerate even major uncomplicated surgery well. The urgency of surgery often precludes otherwise indicated preoperative SNS. Early postoperative PN is usually indicated for these patients, for they are at increased risk of postoperative complications and are unlikely to consume an adequate amount of food voluntarily over the next many days. Patients with only mild muscle atrophy, no systemic inflammation, and no postoperative complications do not require postoperative PN unless (1) adequate feeding by mouth has not been achieved by day 5–7 after surgery or (2) there are indications that voluntary feeding will be further delayed. Perioperative immune-enhancing EN reduces morbidity in patients undergoing major elective gastrointestinal surgery.

Cancer SNS plays a crucial role in cancer therapy. Many malignant neoplasms (especially those that involve the gastrointestinal tract or induce systemic inflammation) and their cytotoxic therapies create the conditions for starvation and commonly lead to SRM or CDM. The prevention or treatment of these starvation diseases will improve patients' quality of life and their tolerance to anticancer therapy. EN and PN are generally not prescribed to patients with advanced cancer for which there is no effective anticancer therapy because the side effects and complications of invasive SNS are not counterbalanced by an improved disease trajectory. In some cases, the disease may be progressing but so slowly that the patient will die of the complications of starvation disease long before they would from the cancer. EN or PN is appropriate for these patients.

Advanced Dementia Optimized voluntary nutrition is the key approach in this situation, and it can be used to deal with problems such as disability and dysphagia in patients who get pleasure from eating. There is no evidence that EN or PN improves quality or length of life in patients who have advanced dementia and show little or no interest in food, and the side effects and complications of EN and PN are unpleasant and sometimes dangerous.

REFEEDING SYNDROME

The refeeding syndrome can occur in patients with adapted SRM during the first week of nutritional repletion if carbohydrate and sodium are introduced too rapidly. Carbohydrate provision stimulates insulin secretion, which, owing to its antinatriuretic effect, expands the ECF volume, especially when excessive sodium is provided. Refeeding edema can be minimized by severely limiting sodium provision and increasing carbohydrate provision slowly. Carbohydrate refeeding may stimulate enough intracellular glucose-6-phosphate and glycogen synthesis to seriously lower serum phosphate concentrations. It also increases the downregulated metabolic rate of patients with adapted SRM and stimulates N retention, new cell synthesis, and cellular rehydration. Phosphorus, potassium, and magnesium deficiencies occur and are dangerous during refeeding; their serum concentrations should be measured frequently, and appropriate supplements provided. Left heart failure may occur in predisposed patients; it has three causes: (1) an abrupt increase of intravascular volume due to the administration of fluids and of glucose, which stimulates insulin-mediated renal sodium retention; (2) increased cardiac demand on an atrophic left ventricle created by an insulin-mediated increase of resting energy expenditure; and (3) myocardial deficiencies of potassium, phosphorus, or magnesium. Cardiac arrhythmias may occur. Acute thiamine deficiency encephalopathy is a devastating preventable complication of refeeding, even with simple glucose infusions.

FURTHER READING

- G F et al: ESPEN guidelines on nutritional support for polymorbid internal medicine patients. *Clin Nutr* 37:336, 2018.
- K J: Nutrition risk screening in the ICU. *Curr Opin Clin Nutr Metab Care* 22:159, 2019.

L KJ et al: Nutrition therapy in critical illness: A review of the literature for clinicians. *Crit Care* 24:35, 2020.

S P et al: Economic evaluation of individualized nutritional support in medical inpatients: Secondary analysis of the EFFORT trial. *Clin Nutr* 25:25, 2020.

S K et al: Pathophysiology of critical illness and role of nutrition. *Nutr Clin Pract* 34:12, 2019.

V Z ARH et al: Nutrition therapy and critical illness: Practical guidance for the ICU, post-ICU, and long-term convalescence phases. *Crit Care* 23:368, 2019.

Y DD et al: Advances in nutrition for the surgical patient. *Curr Probl Surg* 56:343, 2019.

Section 3 Liver and Biliary Tract Disease

336

Approach to the Patient with Liver Disease

Marc G. Ghany, Jay H. Hoofnagle



A diagnosis of liver disease usually can be made accurately by careful elicitation of the patient's history, physical examination, and application of a few laboratory tests. In some circumstances, radiologic examinations are helpful or, indeed, diagnostic. Liver biopsy is considered the criterion standard in evaluation of liver disease but is now needed less for diagnosis than for grading (activity) and staging (fibrosis) of disease. Noninvasive means of assessing fibrosis stage have become increasingly helpful and may allow for avoidance of biopsy in a proportion of patients. This chapter provides an introduction to diagnosis and management of liver disease, briefly reviewing the structure and function of the liver; the major clinical manifestations of liver disease; and the use of clinical history, physical examination, laboratory tests, imaging studies, and liver biopsy.

LIVER STRUCTURE AND FUNCTION

The liver is the largest organ of the body, weighing 1–1.5 kg and representing 1.5–2.5% of the lean body mass. The size and shape of the liver vary and generally match the general body shape—long and lean or squat and square. This organ is located in the right upper quadrant of the abdomen under the right lower rib cage against the diaphragm and projects for a variable extent into the left upper quadrant. It is held in place by ligamentous attachments to the diaphragm, peritoneum, great vessels, and upper gastrointestinal organs. The liver receives a dual blood supply; ~20% of the blood flow is oxygen-rich blood from the hepatic artery, and 80% is nutrient-rich blood from the portal vein arising from the stomach, intestines, pancreas, and spleen.

The majority of cells in the liver are hepatocytes, which constitute two-thirds of the organ's mass. The remaining cell types are Kupffer cells (members of the reticuloendothelial system), stellate (Ito or fat-storing) cells, endothelial and blood vessel cells, bile ductular cells, and cells of supporting structures. Viewed by light microscopy, the liver appears to be organized in lobules, with portal areas at the periphery and central veins in the center of each lobule. However, from a functional point of view, the liver is organized into acini, with both hepatic arterial and portal venous blood entering the acinus from the portal areas (zone 1) and then flowing through the sinusoids to the terminal hepatic veins (zone 3); the intervening hepatocytes constitute zone 2. The advantage of viewing the acinus as the physiologic unit of the liver is that this perspective helps to explain the morphologic patterns and zonality of many vascular and biliary diseases not explained by the lobular arrangement.

Portal areas of the liver consist of small veins, arteries, bile ducts, and lymphatics organized in a loose stroma of supporting matrix and small amounts of collagen. Blood flowing into the portal areas is distributed through the sinusoids, passing from zone 1 to zone 3 of the acinus and draining into the terminal hepatic veins (“central veins”). Secreted bile flows in the opposite direction—that is, in a countercurrent pattern from zone 3 to zone 1. The sinusoids are lined by unique endothelial cells that have prominent fenestrae of variable sizes, allowing the free flow of plasma but not of cellular elements. The plasma is thus in direct contact with hepatocytes in the subendothelial space of Disse.

Hepatocytes have distinct polarity. The basolateral side of the hepatocyte lines the space of Disse and is richly lined with microvilli; it exhibits endocytic and pinocytotic activity, with passive and active uptake of nutrients, proteins, and other molecules. The apical pole of the hepatocyte forms the canalicular membranes through which bile components are secreted. The canaliculi of hepatocytes form a fine network, which fuses into the bile ductular elements near the portal areas. Kupffer cells usually lie within the sinusoidal vascular space and represent the largest group of fixed macrophages in the body. The stellate cells are located in the space of Disse but are not usually prominent unless activated, when they produce collagen and matrix. Red blood cells stay in the sinusoidal space as blood flows through the lobules, but white blood cells can migrate through or around endothelial cells into the space of Disse and from there to portal areas, where they can return to the circulation through lymphatics.

Hepatocytes perform numerous and vital roles in maintaining homeostasis and health. These functions include the synthesis of most essential serum proteins (albumin, carrier proteins, coagulation factors, many hormonal and growth factors), the production of bile and its carriers (bile acids, cholesterol, lecithin, phospholipids), the regulation of nutrients (glucose, glycogen, lipids, cholesterol, amino acids), and the metabolism and conjugation of lipophilic compounds (bilirubin, anions, cations, drugs) for excretion in the bile or urine. Measurement of these activities to assess liver function is complicated by the multiplicity and variability of these functions. The most commonly used liver “function” tests are measurements of serum bilirubin, serum albumin, and prothrombin time. The serum bilirubin level is a measure of hepatic conjugation and excretion; the serum albumin level and prothrombin time are measures of protein synthesis. Abnormalities of bilirubin, albumin, and prothrombin time are typical of hepatic dysfunction. Frank liver failure is incompatible with life, and the functions of the liver are too complex and diverse to be subserved by a mechanical pump; a dialysis membrane; or a concoction of infused hormones, proteins, and growth factors.

LIVER DISEASES

While there are many causes of liver disease (Table 336-1), these disorders generally present clinically in a few distinct patterns and are usually classified as hepatocellular, cholestatic (obstructive), or mixed. In *hepatocellular diseases* (such as viral hepatitis and alcoholic liver disease), features of liver injury, inflammation, and necrosis predominate. In *cholestatic diseases*, such as gallstone or malignant obstruction, primary biliary cholangitis (previously referred to as primary biliary cirrhosis), and some drug-induced liver diseases, features of inhibition of bile flow predominate. In a mixed pattern, features of both hepatocellular and cholestatic injury are present (such as in cholestatic forms of viral hepatitis and many drug-induced liver diseases). The pattern of onset and prominence of symptoms can rapidly suggest a diagnosis, particularly if major risk factors are considered, such as the age and sex of the patient and a history of exposure or risk behaviors.

Typical presenting symptoms of liver disease include jaundice, fatigue, itching, right-upper-quadrant pain, nausea, poor appetite, abdominal distention, and intestinal bleeding. At present, however, many patients are diagnosed with liver disease who have no symptoms and who have been found to have abnormalities in biochemical liver tests as a part of a routine physical examination or screening for blood donation or for insurance or employment. The wide availability of batteries of liver tests makes it relatively simple to demonstrate the

TABLE 336-1 Liver Diseases

Inherited hyperbilirubinemia	Liver involvement in systemic diseases
Gilbert syndrome	Sarcoidosis
Crigler-Najjar syndrome, types I and II	Amyloidosis
Dubin-Johnson syndrome	Glycogen storage diseases
Rotor syndrome	Celiac disease
Viral hepatitis	Tuberculosis
Hepatitis A	<i>Mycobacterium avium-intracellulare</i> infection
Hepatitis B	Cholestatic syndromes
Hepatitis C	Benign postoperative cholestasis
Hepatitis D	Jaundice of sepsis
Hepatitis E	Total parenteral nutrition-induced jaundice
Others (Epstein-Barr virus [mononucleosis] herpesvirus, cytomegalovirus, adenovirus hepatitis)	Cholestasis of pregnancy
Cryptogenic hepatitis	Cholangitis and cholecystitis
Immune and autoimmune liver diseases	Extrahepatic biliary obstruction (stone, stricture, cancer)
Primary biliary cholangitis	Biliary atresia
Autoimmune hepatitis	Caroli disease
Sclerosing cholangitis	Cryptosporidiosis
Overlap syndromes	Drug-induced liver disease
Graft-versus-host disease	Hepatocellular patterns (isoniazid, acetaminophen)
Allograft rejection	Cholestatic patterns (methyltestosterone)
Genetic liver diseases	Mixed patterns (sulfonamides, phenytoin)
α_1 Antitrypsin deficiency	Micro- and macrovesicular steatosis (methotrexate, fialuridine)
Hemochromatosis	Vascular injury
Wilson disease	Sinusoidal obstruction syndrome
Benign recurrent intrahepatic cholestasis	Budd-Chiari syndrome
Progressive familial intrahepatic cholestasis, types I–III	Ischemic hepatitis
Others (galactosemia, tyrosinemia, cystic fibrosis, Niemann-Pick disease, Gaucher's disease)	Passive congestion
Alcoholic liver disease	Portal vein thrombosis
Acute fatty liver	Nodular regenerative hyperplasia
Acute alcoholic hepatitis	Mass lesions
Laënnec cirrhosis	Hepatocellular carcinoma
Nonalcoholic fatty liver	Cholangiocarcinoma
Steatosis	Adenoma
Steatohepatitis	Focal nodular hyperplasia
Acute fatty liver of pregnancy	Metastatic tumors
	Abscess
	Cysts
	Hemangioma

presence of liver injury as well as to rule it out in someone in whom liver disease is suspected.

Evaluation of patients with liver disease should be directed at (1) establishing the etiologic diagnosis, (2) estimating disease severity (*grading*), and (3) establishing the disease stage (*staging*). *Diagnosis* should focus on the category of disease (hepatocellular, cholestatic, or mixed injury) as well as on the specific etiologic diagnosis. *Grading* refers to assessment of the severity or activity of disease—active or inactive as well as mild, moderate, or severe. *Staging* refers to estimation of the point in the course of the natural history of the disease, whether early or late; or precirrhotic, cirrhotic, or end-stage. This chapter introduces general, salient concepts in the evaluation of patients with liver disease that help lead to the diagnoses discussed in subsequent chapters.

CLINICAL HISTORY

The clinical history should focus on the symptoms of liver disease—their nature, patterns of onset, and progression—and on potential risk factors for liver disease. The manifestations of liver disease include

constitutional symptoms such as fatigue, weakness, nausea, poor appetite, and malaise and the more liver-specific symptoms of jaundice, dark urine, light stools, itching, abdominal pain, and bloating. Symptoms can also suggest the presence of cirrhosis, end-stage liver disease, or complications of cirrhosis such as portal hypertension. Generally, the constellation of symptoms and their patterns of onset, rather than a specific symptom, point to an etiology.

Fatigue is the most common and most characteristic symptom of liver disease. It is variously described as lethargy, weakness, listlessness, malaise, increased need for sleep, lack of stamina, and poor energy. The fatigue of liver disease typically arises after activity or exercise and is rarely present or severe after adequate rest; that is, it is "afternoon" rather than "morning" fatigue. Fatigue in liver disease is often intermittent and variable in severity from hour to hour and day to day. In some patients, it may not be clear whether fatigue is due to the liver disease or due to other problems such as stress, anxiety, sleep disturbance, or a concurrent illness.

Nausea occurs with more severe liver disease and may accompany fatigue or be provoked by smelling food odors or eating fatty foods. Vomiting can occur but is rarely persistent or prominent. Poor appetite with weight loss occurs frequently in acute liver disease but is rare in chronic disease except when cirrhosis is present and advanced. Diarrhea is uncommon in liver disease except with severe jaundice, in which a lack of bile acids reaching the intestine can lead to steatorrhea.

Right-upper-quadrant discomfort or ache ("liver pain") occurs in many liver diseases and is usually marked by tenderness over the liver area. The pain arises from stretching or irritation of Glisson's capsule, which surrounds the liver and is rich in nerve endings. Severe pain is most typical of gallbladder disease, liver abscess, and severe sinusoidal obstruction syndrome (previously known as veno-occlusive disease) but is also an occasional accompaniment of acute hepatitis.

Itching occurs with acute liver disease, appearing early in obstructive jaundice (from biliary obstruction) or drug-induced cholestasis and somewhat later in hepatocellular disease (acute hepatitis). Itching also occurs in chronic liver diseases—typically the cholestatic forms such as primary biliary cholangitis and sclerosing cholangitis, in which it is often the presenting symptom, preceding the onset of jaundice. However, itching can occur in any liver disease, particularly once cirrhosis develops.

Jaundice is the hallmark symptom of liver disease and perhaps the most reliable marker of severity. Patients usually report darkening of the urine before they notice scleral icterus. Jaundice is rarely detectable with a bilirubin level <43 µmol/L (2.5 mg/dL). With severe cholestasis, there will also be lightening of the color of the stools and steatorrhea. Jaundice without dark urine usually indicates indirect (unconjugated) hyperbilirubinemia and is typical of hemolytic anemia and the genetic disorders of bilirubin conjugation, the common and benign form being Gilbert syndrome and the rare and severe form being Crigler-Najjar syndrome. Gilbert syndrome affects up to 5% of the general population; the jaundice in this condition is more noticeable after fasting and with stress.

Major risk factors for liver disease that should be sought in the clinical history include details of alcohol use, medication use (including herbal compounds, birth control pills, and over-the-counter medications), personal habits, sexual activity, travel, exposure to jaundiced or other high-risk persons, injection drug use, recent surgery, remote or recent transfusion of blood or blood products, occupation, accidental exposure to blood or needlestick, and familial history of liver disease.

For assessing the risk of viral hepatitis, a careful history of sexual activity is of particular importance and should include the number of lifetime sexual partners and, for men, a history of having sex with men. Sexual exposure is a common mode of spread of hepatitis B and D but is uncommon for hepatitis C. A family history of hepatitis, liver disease, and liver cancer is also important. Maternal-infant transmission occurs with both hepatitis B and C. Vertical spread of hepatitis B can now be prevented by passive and active immunization of the infant at birth. Additionally, antiviral therapy during the third trimester of pregnancy is now recommended for mothers with levels of HBV DNA >200,000 IU/mL. Vertical spread of hepatitis C is uncommon, but there

are no reliable means of prevention. Transmission is more common among HIV-co-infected mothers and is also linked to prolonged and difficult labor and delivery, early rupture of membranes, internal fetal monitoring, and a high maternal viral load. A history of injection drug use, even in the remote past, is of great importance in assessing the risk for hepatitis B and C. Injection drug use is now the single most common risk factor for hepatitis C. Transfusion with blood or blood products is no longer an important risk factor for acute viral hepatitis. However, blood transfusions received before the introduction of sensitive enzyme immunoassays for antibody to hepatitis C virus in 1992 is an important risk factor for chronic hepatitis C. Blood transfusion before 1986, when screening for antibody to hepatitis B core antigen was introduced, is also a risk factor for hepatitis B. Travel to a developing area of the world, exposure to persons with jaundice, and exposure to young children in day-care centers are risk factors for hepatitis A. Tattooing and body piercing (for hepatitis B and C) and eating shellfish (for hepatitis A) are frequently mentioned but are actually types of exposure that rarely lead to the acquisition of hepatitis.

Hepatitis E is one of the more common causes of jaundice in Asia and Africa but is uncommon in developed nations. In endemic areas, transmission is usually through exposure to fecally contaminated water. Recently, non-travel-related (*autochthonous*) cases of hepatitis E have been described in developed countries, including the United States. These cases appear to be due to strains of hepatitis E virus that are endemic in swine and some wild animals (genotypes 3 and 4). While occasional cases are associated with eating raw or undercooked pork or game (deer and wild boars), most cases of hepatitis E occur without known exposure, predominantly in elderly men without typical risk factors for viral hepatitis. Hepatitis E infection can become chronic in immunosuppressed individuals (such as transplant recipients, patients receiving chemotherapy, or patients with HIV infection), in whom it presents with abnormal serum enzymes in the absence of markers of hepatitis B or C.

A history of alcohol intake is important in assessing the cause of liver disease and in planning management and recommendations. In the United States, for example, at least 70% of adults drink alcohol to some degree, but significant alcohol intake is less common; in population-based surveys, only 5% of individuals have more than two drinks per day, the average drink representing 11–15 g of alcohol. Alcohol consumption associated with an increased rate of alcoholic liver disease is probably more than two drinks (22–30 g) per day in women and three drinks (33–45 g) in men. Most patients with alcoholic cirrhosis have a much higher daily intake and have drunk excessively for ≥10 years before onset of liver disease. In assessing alcohol intake, the history should also focus on whether alcohol abuse or dependence is present. Alcoholism is usually defined by the behavioral patterns and consequences of alcohol intake, not by the amount. *Abuse* is defined by a repetitive pattern of drinking alcohol that has adverse effects on social, family, occupational, or health status. *Dependence* is defined by alcohol-seeking behavior, despite its adverse effects. Many alcoholics demonstrate both dependence and abuse, and dependence is considered the more serious and advanced form of alcoholism. A clinically helpful approach to diagnosis of alcohol dependence and abuse is the use of the CAGE questionnaire (Table 336-2), which is recommended for all medical history-taking.

Family history can be helpful in assessing liver disease. Familial causes of liver disease include Wilson disease; hemochromatosis and

TABLE 336-2 CAGE Questions^a

ACRONYM	QUESTION
C	Have you ever felt you ought to cut down on your drinking?
A	Have people annoyed you by criticizing your drinking?
G	Have you ever felt guilty or bad about your drinking?
E	Have you ever had a drink first thing in the morning to steady your nerves or get rid of a hangover (eye-opener)?

^aOne "yes" response should raise suspicion of an alcohol use problem, and more than one is a strong indication of abuse or dependence.

α_1 antitrypsin deficiency; and the less common inherited pediatric liver diseases—that is, familial intrahepatic cholestasis, benign recurrent intrahepatic cholestasis, and Alagille syndrome. Onset of severe liver disease in childhood or adolescence in conjunction with a family history of liver disease or neuropsychiatric disturbance should lead to investigation for Wilson disease. A family history of cirrhosis, diabetes, or endocrine failure and the appearance of liver disease in adulthood suggest hemochromatosis and should prompt investigation of iron status. Abnormal iron studies in adult patients warrant genotyping of the *HFE* gene for the C282Y and H63D mutations typical of genetic hemochromatosis. In children and adolescents with iron overload, other non-*HFE* causes of hemochromatosis should be sought. A family history of emphysema should lead to investigation of α_1 antitrypsin levels and, if levels are low, for protease inhibitor (Pi) genotype.

■ PHYSICAL EXAMINATION

The physical examination rarely uncovers evidence of liver dysfunction in a patient without symptoms or laboratory findings, nor are most signs of liver disease specific to one diagnosis. Thus, the physical examination complements rather than replaces the need for other diagnostic approaches. In many patients, the physical examination is normal unless the disease is acute or severe and advanced. Nevertheless, the physical examination is important in that it can yield the first evidence of hepatic failure, portal hypertension, and liver decompensation. In addition, the physical examination can reveal signs—related either to risk factors or to associated diseases or findings—that point to a specific diagnosis.

Typical physical findings in liver disease are icterus, hepatomegaly, hepatic tenderness, splenomegaly, spider angioma, palmar erythema, and skin excoriations. Signs of advanced disease include muscle wasting, ascites, edema, dilated abdominal veins, hepatic fetor, asterixis, mental confusion, stupor, and coma. In male patients with cirrhosis, particularly that related to alcohol use, signs of hyperestrogenemia such as gynecomastia, testicular atrophy, and loss of male-pattern hair distribution may be found.

Icterus is best appreciated when the sclera is inspected under natural light. In fair-skinned individuals, a yellow tinge to the skin may be obvious. In dark-skinned individuals, examination of the mucous membranes below the tongue can demonstrate jaundice. Jaundice is rarely detectable if the serum bilirubin level is $<43 \mu\text{mol/L}$ (2.5 mg/dL) but may remain detectable below this level during recovery from jaundice (because of protein and tissue binding of conjugated bilirubin).

Spider angioma and palmar erythema occur in both acute and chronic liver disease; these manifestations may be especially prominent in persons with cirrhosis but can develop in normal individuals and are frequently found during pregnancy. Spider angioma are superficial, tortuous arterioles, and—unlike simple telangiectasias—typically fill from the center outward. Spider angioma occur only on the arms, face, and upper torso; they can be pulsatile and may be difficult to detect in dark-skinned individuals.

Hepatomegaly is not a very reliable sign of liver disease because of variability in the liver's size and shape and the physical impediments to assessment of liver size by percussion and palpation. Marked hepatomegaly is typical of cirrhosis, sinusoidal obstruction syndrome, infiltrative disorders such as amyloidosis, metastatic or primary cancers of the liver, and alcoholic hepatitis. Careful assessment of the liver edge may also reveal unusual firmness, irregularity of the surface, or frank nodules. Perhaps the most reliable physical finding in the liver examination is hepatic tenderness. Discomfort when the liver is touched or pressed upon should be carefully sought with percussive comparison of the right and left upper quadrants.

Splenomegaly, which occurs in many medical conditions, can be a subtle but significant physical finding in liver disease. The availability of ultrasound (US) methods for assessment of the spleen allows confirmation of the physical finding.

Signs of advanced liver disease include muscle wasting and weight loss as well as hepatomegaly, bruising, ascites, and edema. Ascites is best appreciated by attempts to detect shifting dullness by careful percussion. US examination will confirm the finding of ascites in

equivocal cases. Peripheral edema can occur with or without ascites. In patients with advanced liver disease, other factors frequently contribute to edema formation, including hypoalbuminemia, venous insufficiency, heart failure, and medications.

Hepatic failure is defined as the occurrence of signs or symptoms of hepatic encephalopathy in a person with severe acute or chronic liver disease. The first signs of hepatic encephalopathy can be subtle and nonspecific—change in sleep patterns, change in personality, irritability, and mental dullness. Thereafter, confusion, disorientation, stupor, and eventually coma supervene. In acute liver failure, excitability and mania may be present. Physical findings include asterixis and flapping tremors of the body and tongue. *Fetor hepaticus* refers to the slightly sweet, ammoniacal odor that can develop in patients with liver failure, particularly if there is portal-venous shunting of blood around the liver. Other causes of coma and disorientation should be excluded, mainly electrolyte imbalances, sedative use, and renal or respiratory failure. The appearance of hepatic encephalopathy during acute hepatitis is the major criterion for diagnosis of fulminant hepatitis and indicates a poor prognosis. In chronic liver disease, encephalopathy is usually triggered by a medical complication such as gastrointestinal bleeding, overdiuresis, uremia, dehydration, electrolyte imbalance, infection, constipation, or use of narcotic analgesics.

A helpful measure of hepatic encephalopathy is a careful mental status examination and use of the trail-making test, which consists of a series of 25 numbered circles that the patient is asked to connect as rapidly as possible using a pencil. The normal range for the connect-the-dot test is 15–30 s; it is considerably longer in patients with early hepatic encephalopathy. Other tests include drawing of abstract objects or comparison of a signature to previous examples. More sophisticated testing—for example, with electroencephalography and visual evoked potentials—can detect mild forms of encephalopathy but are rarely clinically useful.

Other signs of advanced liver disease include umbilical hernia from ascites, hydrothorax, prominent veins over the abdomen, and *caput medusa*, a condition that consists of collateral veins radiating from the umbilicus and results from recanalization of the umbilical vein. Widened pulse pressure and signs of a hyperdynamic circulation can occur in patients with cirrhosis as a result of fluid and sodium retention, increased cardiac output, and reduced peripheral resistance. Patients with long-standing cirrhosis and portal hypertension are prone to develop the hepatopulmonary syndrome, which is defined by the triad of liver disease, hypoxemia, and pulmonary arteriovenous shunting. The hepatopulmonary syndrome is characterized by platypnea and orthodeoxia: shortness of breath and oxygen desaturation that occur paradoxically upon the assumption of an upright position. Measurement of oxygen saturation by pulse oximetry is a reliable screening test for hepatopulmonary syndrome.

Several skin disorders and changes are common in liver disease. Hyperpigmentation is typical of advanced chronic cholestatic diseases such as primary biliary cholangitis and sclerosing cholangitis. In these same conditions, xanthelasma and tendon xanthomata occur as a result of retention and high serum levels of lipids and cholesterol. Slate-gray pigmentation of the skin is also seen with hemochromatosis if iron levels are high for a prolonged period. Mucocutaneous vasculitis with palpable purpura, especially on the lower extremities, is typical of cryoglobulinemia of chronic hepatitis C but can also occur in chronic hepatitis B.

Some physical signs point to specific liver diseases. Kayser-Fleischer rings occur in Wilson disease and consist of a golden-brown copper pigment deposited in Descemet's membrane at the periphery of the cornea; they are best seen by slit-lamp examination. Dupuytren contracture and parotid enlargement are suggestive of chronic alcoholism and alcoholic liver disease. In metastatic liver disease or primary hepatocellular carcinoma, signs of cachexia and wasting as well as firm hepatomegaly and a hepatic bruit may be prominent.

■ DIAGNOSIS OF LIVER DISEASE

The key diagnostic tests of major causes of acute and chronic liver disease are outlined in Table 336-3, and an algorithm for evaluation of

TABLE 336-3 Important Diagnostic Tests in Common Liver Diseases

DISEASE	DIAGNOSTIC TEST
Hepatitis A	Anti-HAV IgM
Hepatitis B	
Acute	HBsAg and anti-HBc IgM
Chronic	HBsAg and HBeAg and/or HBV DNA
Hepatitis C	Anti-HCV and HCV RNA
Hepatitis D (delta)	HBsAg and anti-HDV
Hepatitis E	Anti-HEV IgM and HEV RNA
Autoimmune hepatitis	ANA or SMA, elevated IgG levels, and compatible histology
Primary biliary cholangitis	Mitochondrial antibody, elevated IgM levels, and compatible histology
Primary sclerosing cholangitis	P-ANCA, cholangiography
Drug-induced liver disease	History of drug ingestion
Alcoholic liver disease	History of excessive alcohol intake and compatible histology
Nonalcoholic steatohepatitis	Ultrasound or CT evidence of fatty liver and compatible histology
α_1 Antitrypsin disease	Reduced α_1 antitrypsin levels, phenotype PiZZ or PiSZ
Wilson disease	Decreased serum ceruloplasmin and increased urinary copper; increased hepatic copper level
Hemochromatosis	Elevated iron saturation and serum ferritin; genetic testing for <i>HFE</i> gene mutations
Hepatocellular cancer	Elevated α -fetoprotein level (to >500 ng/mL); ultrasound or CT image of mass

Abbreviations: ANA, antinuclear antibody; anti-HBc, antibody to hepatitis B core antigen; HAV, HBV, HCV, HDV, HEV, hepatitis A, B, C, D, E virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; P-ANCA, peripheral antineutrophil cytoplasmic antibody; SMA, smooth-muscle antibody.

the patient with suspected liver disease is shown in Fig. 336-1. Specifics of diagnosis are discussed in later chapters. The most common causes of acute liver disease are viral hepatitis (particularly hepatitis A, B, and C), drug-induced liver injury, cholangitis, and alcoholic liver disease. Liver biopsy usually is not needed for the diagnosis and management of acute liver disease, exceptions being situations where the diagnosis remains unclear despite thorough clinical and laboratory investigation. Liver biopsy can be helpful in diagnosing drug-induced liver disease and acute alcoholic hepatitis.

The most common causes of chronic liver disease, in general order of frequency, are chronic hepatitis C, alcoholic liver disease, nonalcoholic steatohepatitis, chronic hepatitis B, autoimmune hepatitis, sclerosing cholangitis, primary biliary cholangitis, hemochromatosis, and Wilson disease. Hepatitis E virus is a rare cause of chronic hepatitis, with cases occurring mostly in persons who are immunosuppressed or immunodeficient. Strict diagnostic criteria have not been developed for most liver diseases, but liver biopsy plays an important role in the diagnosis of autoimmune hepatitis, primary biliary cholangitis, nonalcoholic and alcoholic steatohepatitis, and Wilson disease (with a quantitative hepatic copper level in the last instance).

Laboratory Testing Diagnosis of liver disease is greatly aided by the availability of reliable and sensitive tests of liver injury and function. A typical battery of blood tests used for initial assessment of liver disease includes measurement of levels of serum alanine (ALT) and aspartate (AST) aminotransferases, alkaline phosphatase (AlkP), direct and total serum bilirubin and albumin, and prothrombin time. The pattern of abnormalities generally points to hepatocellular versus cholestatic liver disease and helps determine whether the disease is acute or chronic and whether cirrhosis and hepatic failure are present. Based on these results, further testing over time may be necessary. Other laboratory tests may be helpful, such as γ -glutamyl transpeptidase (γ GT) to define whether AlkP elevations are due to liver disease; hepatitis serology to define the type of viral hepatitis; and autoimmune markers to diagnose primary biliary cholangitis (antimitochondrial

antibody), sclerosing cholangitis (peripheral antineutrophil cytoplasmic antibody), and autoimmune hepatitis (antinuclear, smooth-muscle, and liver-kidney microsomal antibody). A simple delineation of laboratory abnormalities and common liver diseases is given in Table 336-3.

The use and interpretation of liver function tests are summarized in Chap. 337.

Diagnostic Imaging Great advances have been made in hepatobiliary imaging, although no method is adequately accurate in demonstrating underlying cirrhosis in its early stages. Of the many modalities available for imaging the liver, US, computed tomography (CT), and magnetic resonance imaging (MRI) are the most commonly employed and are complementary to one another. In general, US and CT are highly sensitive for detecting biliary duct dilation and are the first-line options for investigating cases of suspected obstructive jaundice. All three modalities can detect a fatty liver, which appears bright on imaging studies. Modifications of CT and MRI can be used to quantify liver fat, and this information may ultimately be valuable in monitoring response to therapy in patients with fatty liver disease. Advantages, disadvantages, and clinical utility of each modality are presented in Table 336-4. Magnetic resonance cholangiopancreatography (MRCP) and endoscopic retrograde cholangiopancreatography (ERCP) are the procedures of choice for visualization of the biliary tree. MRCP offers several advantages over ERCP: there is no need for contrast media or ionizing radiation, images can be acquired faster, the procedure is less operator dependent, and it carries no risk of pancreatitis. MRCP is superior to US and CT for detecting choledocholithiasis but is less specific. MRCP is useful in the diagnosis of bile duct obstruction and congenital biliary abnormalities, but ERCP is considered more valuable in evaluating ampullary lesions and primary sclerosing cholangitis. ERCP permits biopsy, direct visualization of the ampulla and common bile duct, and intraductal ultrasonography and brushings for cytologic evaluation of malignancy. It also provides several therapeutic options in patients with obstructive jaundice, such as sphincterotomy, stone extraction, and placement of nasobiliary catheters and biliary stents.

Doppler US and MRI are used to assess hepatic vasculature and hemodynamics and to monitor surgically or radiologically placed vascular shunts, including transjugular intrahepatic portosystemic shunts. Multidetector or spiral CT and MRI with contrast enhancement are the procedures of choice for the identification and evaluation of hepatic masses, the staging of liver tumors, and preoperative assessment. With regard to mass lesions, the sensitivity of hepatic imaging continues to increase; unfortunately, specificity remains a problem, and often two and sometimes three studies are needed before a diagnosis can be reached. An emerging imaging modality for the investigation of hepatic lesions is contrast-enhanced US. This procedure permits enhancement of liver lesions in a similar fashion as contrast-enhanced, cross-sectional CT or MRI. Major advantages are real-time assessment of liver perfusion throughout the vascular phases without risk of nephrotoxicity and radiation exposure. Other advantages are its widespread availability and lower cost. Limitations include body habitus of the patient and skill of the operator. US is the recommended modality for hepatocellular carcinoma (HCC) screening. Contrast-enhanced US, CT, and MRI are appropriate for further investigation of lesions detected on screening US. The American College of Radiologists has developed a Liver Imaging Reporting and Data System (LI-RADS) to standardize the reporting and data collection of CT, MRI, and contrast-enhanced US imaging for HCC. This system allows for more consistent reporting and reduces imaging interpretation variability and errors.

Recently, several US-based elastographic techniques have been developed and approved for the measurement of hepatic stiffness, providing an indirect assessment of fibrosis and cirrhosis. The most commonly used approaches in clinical practice include transient elastography, acoustic radiation force impulse imaging, shear-wave elasticity imaging, and supersonic shear imaging. These techniques can eliminate the need for liver biopsy if the only indication for the test is the assessment of disease stage. Magnetic resonance elastography is

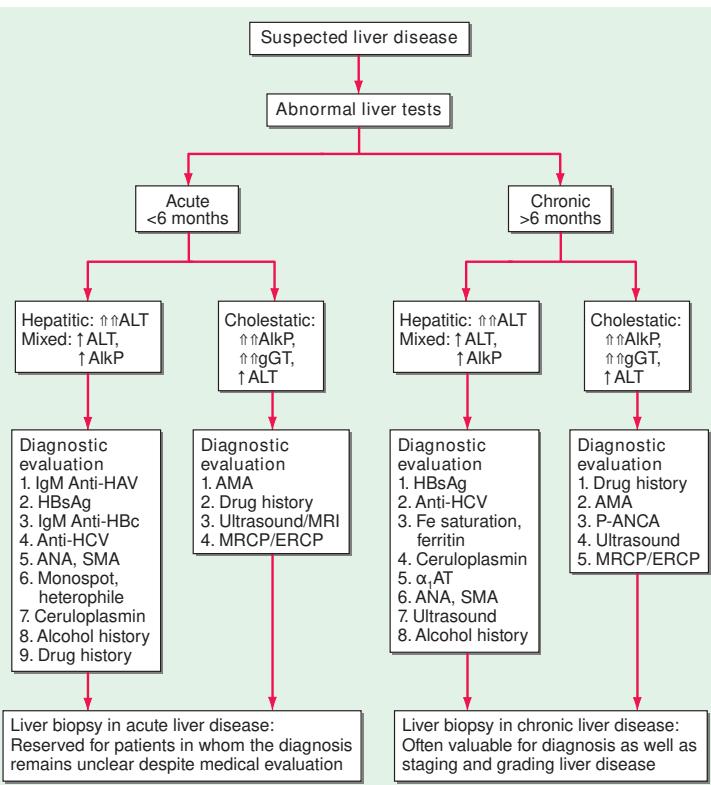


FIGURE 336-1 Algorithm for evaluation of abnormal liver tests. For patients with suspected liver disease, an appropriate approach to evaluation is initial routine liver testing—for example, measurement of serum bilirubin, albumin, alanine aminotransferase (ALT), AST, and alkaline phosphatase (AlkP). These results (sometimes complemented by testing of γ -glutamyl transpeptidase [gGT]) will establish whether the pattern of abnormalities is hepatic, cholestatic, or mixed. In addition, the duration of symptoms or abnormalities will indicate whether the disease is acute or chronic. If the disease is acute and if history, laboratory tests, and imaging studies do not reveal a diagnosis, liver biopsy is appropriate to help establish the diagnosis. If the disease is chronic, liver biopsy can be helpful not only for diagnosis but also for grading of the activity and staging the progression of disease. This approach is generally applicable to patients without immune deficiency. In patients with HIV infection or recipients of bone marrow or solid organ transplants, the diagnostic evaluation should also include evaluation for opportunistic infections (e.g., with adenovirus, cytomegalovirus, *Coccidioides*, hepatitis E virus) as well as for vascular and immunologic conditions (veno-occlusive disease, graft-versus-host disease), α , AT, α , antitrypsin; AMA: antimitochondrial antibody; ANA, antinuclear antibody; anti-HBc, antibody to hepatitis B core (antigen); ERCP, endoscopic retrograde cholangiopancreatography; HAV, hepatitis A virus; HBsAg, hepatitis B surface antigen; HCV, hepatitis C virus; MRCP, magnetic resonance cholangiopancreatography; P-ANCA, peripheral antineutrophil cytoplasmic antibody; SMA, smooth-muscle antibody.

more sensitive than US elastography but is also more expensive and requires advanced scheduling and special equipment. Studies are ongoing to determine whether hepatic elastography is an appropriate means of monitoring fibrosis and disease progression in untreated and treated patients. Finally, interventional radiologic techniques allow for the biopsy of solitary lesions, the radiofrequency ablation and chemoembolization of cancerous lesions, the insertion of drains into hepatic abscesses, the measurement of portal pressure, and the creation of vascular shunts in patients with portal hypertension. Which modality to use depends on factors such as availability, cost, and experience of the radiologist with each technique.

Liver Biopsy Liver biopsy remains the gold standard in the evaluation of patients with liver disease, particularly chronic liver disease. Liver biopsy is necessary for diagnosis in selected instances but is more often useful for assessment of the severity (grade) and stage of liver damage, prediction of prognosis, and monitoring of the response to treatment. The size of the liver biopsy sample is an important determinant of reliability; a length of 1.5–2 cm with 10 portal tracts is necessary for accurate assessment of fibrosis. Because liver biopsy is an invasive procedure and not without complications, it should be used only when it will contribute materially to decisions about management and therapy. In the future, noninvasive means of assessing disease activity (batteries of blood tests) and fibrosis (elastography and fibrosis markers) may replace liver biopsy for the staging and grading of disease.

■ GRADING AND STAGING OF LIVER DISEASE

Grading refers to an assessment of the severity or activity of liver disease, whether acute or chronic; active or inactive; and mild, moderate, or severe. Liver biopsy is the most accurate means of assessing severity, particularly in chronic liver disease. Serum aminotransferase levels serve as convenient and noninvasive markers for disease activity but do not always reliably reflect disease severity. Thus, normal serum aminotransferase levels in patients with hepatitis B surface antigen in serum may indicate the inactive carrier state or may reflect mild chronic hepatitis B or hepatitis B with fluctuating disease activity. Serum testing for hepatitis B e antigen and hepatitis B virus DNA can help sort out these different patterns, but these markers can also fluctuate and change

TABLE 336-4 Diagnostic Tests to Assess Liver Fat

IMAGING MODALITY	ADVANTAGES	DISADVANTAGES	CLINICAL UTILITY
Ultrasound	No radiation Widely available	Operator dependent Imprecise qualitative assessment of fat severity, particularly mild steatosis	Initial screening test for suspected liver fat
Transient elastography with controlled attenuation parameter	No radiation Point-of-care assessment of liver fat Provides semiquantitative assessment of fat severity	Requires special software No reliable cutoff for diagnosis of liver fat Imprecise qualitative assessment of fat severity	Alternate screening test for suspected liver fat if available
Computed tomography	Rapid assessment Non-operator dependent Quantitative assessment of fat severity	Requires radiation Quantification of fat requires specific protocols Imprecise quantitative assessment of fat severity, particularly mild steatosis	Not recommended for clinical assessment of liver fat due to need for radiation exposure and low sensitivity for mild fat
Magnetic resonance imaging—proton density fat fraction	Direct assessment of liver fat Highly sensitive and specific	Relatively limited accessibility	Test of choice for quantitative assessment of liver fat if available

over time. Similarly, in chronic hepatitis C, serum aminotransferase levels can be normal despite moderate disease activity. Finally, in both alcoholic and nonalcoholic steatohepatitis, aminotransferase levels are quite unreliable in reflecting severity. In these conditions, liver biopsy is helpful in guiding management and identifying appropriate therapy, particularly if treatment is difficult, prolonged, and expensive, as is often the case in chronic viral hepatitis. Of the several well-verified numerical scales for grading activity in chronic liver disease, the most commonly used are the METAVIR, histology activity index, and the Ishak fibrosis scale.

Liver biopsy is also the most accurate means of assessing stage of disease as early or advanced, precirrhotic, and cirrhotic. Staging of disease pertains largely to chronic liver diseases in which progression to cirrhosis and end-stage disease can occur but may require years or decades. Clinical features, biochemical tests, and hepatic imaging studies are helpful in assessing stage but generally become abnormal only in the middle to late stages of cirrhosis. Noninvasive tests that suggest advanced fibrosis include mild elevations of bilirubin, prolongation of prothrombin time, slight decreases in serum albumin, and mild thrombocytopenia (which is often the first indication of worsening fibrosis). Combinations of blood test results that include clinical features, routine laboratory tests, and special laboratory tests such as serum proteins or small molecules that are affected by or involved with fibrogenesis have been used to create models for predicting advanced liver disease, but these models are not reliable enough to use on a regular basis or for repeated measures and only separate advanced from early disease (Table 336-5). Recently, elastography and noninvasive breath tests using ¹³C-labeled compounds have been proposed as a means of detecting early stages of fibrosis and liver dysfunction, but their reliability and reproducibility remain to be proven. A major limitation of noninvasive markers is that they can be affected by disease activity. Even elastography is limited in this regard, in that it measures liver stiffness, not fibrosis per se, and can be affected by inflammation, edema, hepatocyte necrosis, and intrasinusoidal cellularity (inflammatory, malignant, or sickled cells). Thus, at present, mild to moderate stages of hepatic fibrosis are detectable only by liver biopsy. In the assessment of stage, the degree of fibrosis is usually used as the quantitative measure. The amount of fibrosis is generally staged on a scale of 0 to 4+ (METAVIR scale) or 0 to 6+ (Ishak scale). The importance of staging relates primarily to prognosis, recommendation of therapy, and optimal management to prevent complications of chronic liver disease. Patients with cirrhosis are candidates for screening and surveillance for esophageal varices and HCC. Patients without advanced fibrosis need not undergo screening.

TABLE 336-5 Selected Noninvasive Methods of Assessing Hepatic Fibrosis and Cirrhosis

METHOD	PARAMETERS	ADVANCED FIBROSIS	CIRRHOSIS
APRI	AST, platelet count	>1	>1.5 (1–2)
ELF	Age, hyaluronic acid, MMP-3, TIMP-1	>7.7	>9.3
FIB-4	Age, AST, ALT, platelet count	>1.45	>3.25
Fibro test ^a	Haptoglobin, α_2 -macroglobulin, apolipoprotein A1, γ GT, total bilirubin	>0.45	>0.63
TE	Measures speed of a shear wave generated by vibration through liver tissue	>7.3 kPa	>15 kPa (9–26.5 kPa)
ARFI	Measures speed of shear wave generated by acoustic radiation force through liver tissue	>1.3 m/s	>1.87 m/s

^aPatented models.

Note: The cut points presented in the table were mostly derived from patients with chronic hepatitis C. The cut points for the noninvasive models and tests presented in the table vary among different liver diseases and among patients with the same disease among different populations.

Abbreviations: ALT, alanine aminotransferase; APRI, AST-to-platelet ratio; ARFI, acoustic radiation force imaging; AST, aspartate aminotransferase; ELF, enhanced liver fibrosis panel; γ GT, γ -glutamyl transpeptidase; MMP-3, metalloproteinase-3; TIMP-1, tissue inhibitor of metalloproteinase-1; TE, transient elastography.

TABLE 336-6 Child-Pugh Classification of Cirrhosis

FACTOR	UNITS	POINTS TOWARD TOTAL SCORE		
		1	2	3
Serum bilirubin	$\mu\text{mol/L}$	<34	34–51	>51
	mg/dL	<2.0	2.0–3.0	>3.0
Serum albumin	g/L	>35	30–35	<30
	g/dL	>3.5	3.0–3.5	<3.0
Prothrombin time	seconds prolonged	<4	4–6	>6
	INR ^a	<1.7	1.7–2.3	>2.3
Ascites		None	Easily controlled	Poorly controlled
Hepatic encephalopathy		None	Minimal	Advanced

^aInternational normalized ratio.

Note: The Child-Pugh score is calculated by adding the scores for the five factors and can range from 5 to 15. The resulting Child-Pugh class can be A (a score of 5–6), B (7–9), or C (\geq 10). Decompensation indicates cirrhosis, with a Child-Pugh score of \geq 7 (class B). This level has been the accepted criterion for listing a patient for liver transplantation.

Once cirrhosis develops, other scoring systems are employed to assess compensated versus decompensated disease and prognosis. The first staging system used for this purpose was the modified Child-Pugh classification, with a scoring system of 5–15: scores of 5 and 6 represent Child-Pugh class A (consistent with “compensated cirrhosis”), scores of 7–9 represent class B, and scores of 10–15 represent class C (Table 336-6). This scoring system was initially devised to stratify patients with cirrhosis into risk groups before portal decompressive surgery. The Child-Pugh score is a reasonably reliable predictor of survival in many liver diseases and predicts the likelihood of major complications of cirrhosis, such as bleeding from varices and spontaneous bacterial peritonitis. This classification scheme was used to assess prognosis in cirrhosis and to provide standard criteria for listing a patient as a candidate for liver transplantation (Child-Pugh class B). More recently, the Child-Pugh system has been replaced by the Model for End-Stage Liver Disease (MELD) system for the latter purpose. The MELD score is a prospectively derived system designed to predict the prognosis of patients with liver disease and portal hypertension. This score is calculated using three readily available objective variables: the prothrombin time expressed as the international normalized ratio (INR), the serum bilirubin level, and the serum creatinine concentration. The ability of the MELD score to predict outcome after liver transplantation is regularly monitored and was modified to increase its accuracy and improve allocation of donated livers. These modifications include serum sodium concentration as a factor in the model and a reweighting of the MELD components. A separate scoring system, the Pediatric End-Stage Liver Disease (PELD) score, is used for children (<12 years old). Transient elastography has also been used to stage cirrhosis and has been shown to be useful in predicting complications such as variceal hemorrhage, ascites development, and liver-related death.

The MELD system provides a more objective means of assessing disease severity and has less center-to-center variation than the Child-Pugh score as well as a wider range of values. The MELD and PELD systems are currently used to establish priority listing for liver transplantation in the United States. Convenient MELD and PELD calculators are available via the Internet (<https://optn.transplant.hrsa.gov/resources/allocation-calculators/about-meld-and-pehd/>).

NONSPECIFIC ISSUES IN THE MANAGEMENT OF PATIENTS WITH LIVER DISEASE

Specifics on the management of different forms of acute or chronic liver disease are supplied in subsequent chapters, but certain issues are applicable to any patient with liver disease. These issues include advice regarding alcohol use, medication use, vaccination, and surveillance for certain liver diseases and complications of liver disease. Alcohol should be used sparingly, if at all, by patients with liver

disease. Abstinence from alcohol should be encouraged for all patients with alcohol-related liver disease, patients with cirrhosis, and patients receiving interferon-based therapy for hepatitis B and during anti-viral therapy of hepatitis C. With regard to vaccinations, all patients with liver disease should receive hepatitis A vaccine, and those with risk factors should receive hepatitis B vaccine as well. Influenza and pneumococcal vaccination should also be encouraged, with adherence to the recommendations of the Centers for Disease Control and Prevention (CDC). Patients with liver disease should exercise caution in using any medications other than those that are most necessary. Drug-induced hepatotoxicity can mimic many forms of liver disease and can cause exacerbations of chronic hepatitis and cirrhosis; drugs should be suspected in any situation in which the cause of exacerbation is unknown. The CDC now recommends universal one-time testing for hepatitis C virus among persons aged 18–79 years and screening of all pregnant women during each pregnancy except in settings where the prevalence of hepatitis C virus infection (hepatitis C virus RNA positivity) is <0.1%. Finally, consideration should be given to surveillance for complications of chronic liver disease such as variceal hemorrhage and HCC. Cirrhosis warrants upper endoscopy to assess the presence of varices, and the patient should receive chronic therapy with beta blockers or should be offered endoscopic obliteration if large varices are found. Moreover, cirrhosis warrants screening and long-term surveillance for development of HCC. While the optimal regimen for such surveillance has not been established, an appropriate approach is US of the liver at 6- to 12-month intervals.

FURTHER READING

- F SL et al: Mechanisms of NAFLD development and therapeutic strategies. *Nat Med* 24:908, 2018.
- S WK et al: Chronic hepatitis B virus infection. *Lancet* 392:2313, 2018.
- S CW et al: Hepatitis C. *Lancet* 394:1451, 2019.
- T EB, L AS: Use of liver imaging and biopsy in clinical practice. *N Engl J Med* 377:756, 2017.

337

Evaluation of Liver Function

Emily D. Bethea, Daniel S. Pratt

There are a number of tests that can be used to evaluate liver function. These tests include biochemical tests, radiologic tests, and pathologic tests.

Serum biochemical tests, also commonly referred to as “liver function tests,” can be used to (1) detect the presence of liver disease, (2) distinguish among different types of liver disorders, (3) gauge the extent of known liver damage, and (4) follow the response to treatment. However, serum biochemical tests have shortcomings. They lack sensitivity and specificity; they can be normal in patients with serious liver disease and abnormal in patients with diseases that do not affect the liver. Liver tests rarely suggest a specific diagnosis; rather, they suggest a general category of liver disease, such as hepatocellular or cholestatic, which then further directs the evaluation. The liver carries out thousands of biochemical functions, most of which cannot be easily measured by blood tests. Laboratory tests measure only a limited number of these functions. In fact, many tests, such as the aminotransferases and alkaline phosphatase, do not measure liver function at all. Rather, they detect liver cell damage or interference with bile flow. Thus, no one biochemical test enables the clinician to accurately assess the liver’s total functional capacity.

To increase the sensitivity and the specificity of biochemical tests in the detection of liver disease, it is best to use them as a battery. Tests usually employed in clinical practice include the bilirubin,

aminotransferases, alkaline phosphatase, albumin, and prothrombin time tests. When more than one of these tests provide abnormal findings or the findings are persistently abnormal on serial determinations, the probability of liver disease is high. When all test results are normal, the probability of missing occult liver disease is low.

Serum Bilirubin (See also Chap. 49) Bilirubin, a breakdown product of the porphyrin ring of heme-containing proteins, is found in the blood in two fractions—conjugated and unconjugated. The unconjugated fraction, also termed the *indirect fraction*, is insoluble in water and is bound to albumin in the blood. The conjugated (direct) bilirubin fraction is water-soluble and can therefore be excreted by the kidney. Normal values of total serum bilirubin are reported between 1 and 1.5 mg/dL with 95% of a normal population falling between 0.2 and 0.9 mg/dL. If the direct-acting fraction is <15% of the total, the bilirubin can be considered to all be indirect. The most frequently reported upper limit of normal for conjugated bilirubin is 0.3 mg/dL.

Elevation of the unconjugated fraction of bilirubin is rarely due to liver disease. An isolated elevation of unconjugated bilirubin is seen primarily in hemolytic disorders and in a number of genetic conditions such as Crigler-Najjar and Gilbert’s syndromes (Chap. 49). Isolated unconjugated hyperbilirubinemia (bilirubin elevated but <15% direct) should prompt a workup for hemolysis (Fig. 337-1). In the absence of hemolysis, an isolated, unconjugated hyperbilirubinemia in an otherwise healthy patient can be attributed to Gilbert’s syndrome, and no further evaluation is required.

In contrast, conjugated hyperbilirubinemia almost always implies liver or biliary tract disease. The rate-limiting step in bilirubin metabolism is not conjugation of bilirubin, but rather the transport of conjugated bilirubin into the bile canaliculi. Thus, elevation of the conjugated fraction may be seen in any type of liver disease including fulminant liver failure. In most liver diseases, both conjugated and unconjugated fractions of the bilirubin tend to be elevated. Except in the presence of a purely unconjugated hyperbilirubinemia, fractionation of the bilirubin is rarely helpful in determining the cause of jaundice.

Although the degree of elevation of the serum bilirubin has not been critically assessed as a prognostic marker, it is important in a number of conditions. In viral hepatitis, the higher the serum bilirubin, the greater is the hepatocellular damage. Total serum bilirubin correlates with poor outcomes in alcoholic hepatitis. It is also a critical component of the Model for End-Stage Liver Disease (MELD) score, a tool used to estimate survival of patients with end-stage liver disease, prioritize patients awaiting liver transplantation, and assess operative risk of patients with cirrhosis. An elevated total serum bilirubin in patients with drug-induced liver disease indicates more severe injury.

Unconjugated bilirubin always binds to albumin in the serum and is not filtered by the kidney. Therefore, any bilirubin found in the urine is conjugated bilirubin; the presence of bilirubinuria implies the presence of liver disease or obstructive jaundice. A urine dipstick test can theoretically give the same information as fractionation of the serum bilirubin. This test is almost 100% accurate. Phenothiazines may give a false-positive reading with the Ictotest tablet. In patients recovering from jaundice, the urine bilirubin clears prior to the serum bilirubin.

Serum Enzymes The liver contains thousands of enzymes, some of which are also present in the serum in very low concentrations. These enzymes have no known function in the serum and behave like other serum proteins. They are distributed in the plasma and in interstitial fluid and have characteristic half-lives, which are usually measured in days. Very little is known about the catabolism of serum enzymes, although they are probably cleared by cells in the reticuloendothelial system. The elevation of a given enzyme activity in the serum is thought to primarily reflect its increased rate of entrance into serum from damaged liver cells.

Serum enzyme tests can be grouped into two categories: (1) enzymes whose elevation in serum reflects damage to hepatocytes and (2) enzymes whose elevation in serum reflects cholestasis.

ENZYMESTHAT REFLECT DAMAGE TO HEPATOCYTES The aminotransferases (transaminases) are sensitive indicators of liver cell injury and

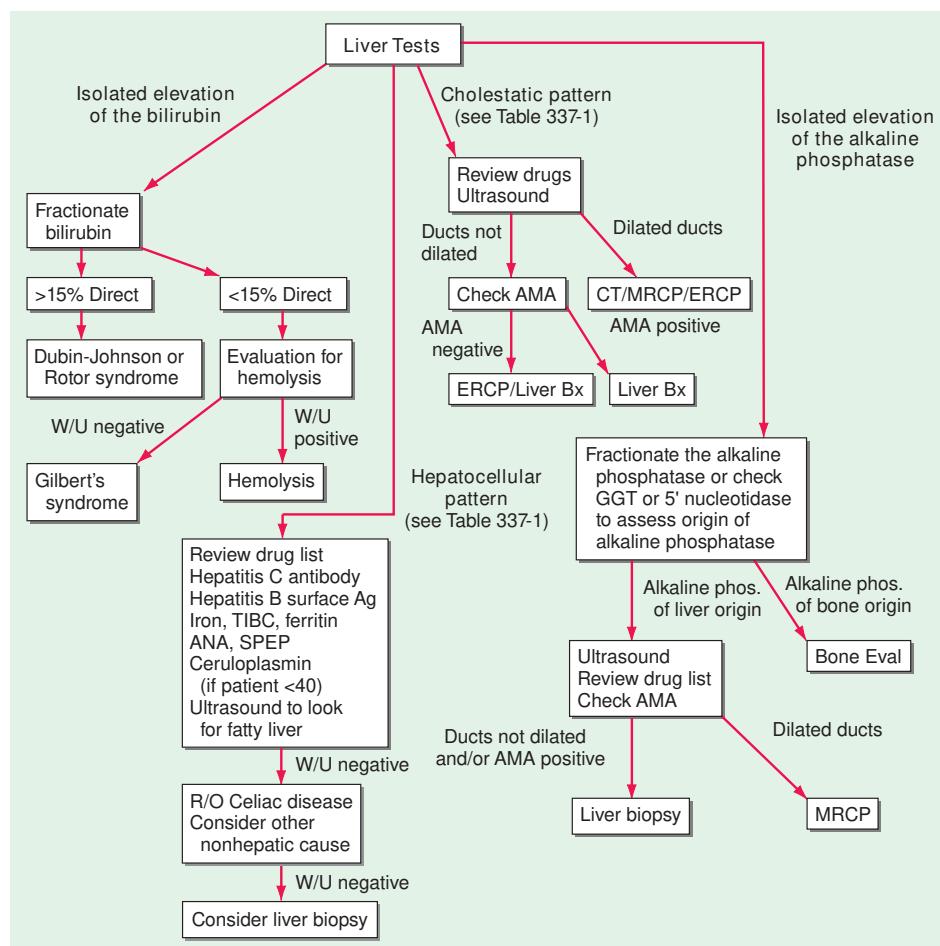


FIGURE 337-1 Algorithm for the evaluation of chronically abnormal liver tests. Ag, antigen; AMA, antimitochondrial antibody; ANA, antinuclear antibody; Bx, biopsy; CT, computed tomography; ERCP, endoscopic retrograde cholangiopancreatography; GGT, γ -glutamyl transpeptidase; MRCP, magnetic resonance cholangiopancreatography; R/O, rule out; SPEP, serum protein electrophoresis; TIBC, total iron-binding capacity; W/U, workup.

are most helpful in recognizing acute hepatocellular diseases such as hepatitis. They include aspartate aminotransferase (AST) and alanine aminotransferase (ALT). AST is found in the liver, cardiac muscle, skeletal muscle, kidneys, brain, pancreas, lungs, leukocytes, and erythrocytes in decreasing order of concentration. ALT is found primarily in the liver and is therefore a more specific indicator of liver injury. The aminotransferases are normally present in the serum in low concentrations. These enzymes are released into the blood in greater amounts when there is damage to the liver cell membrane, resulting in increased permeability. Liver cell necrosis is not required for the release of the aminotransferases, and there is a poor correlation between the degree of liver cell damage and the level of the aminotransferases. Thus, the absolute elevation of the aminotransferases is of no prognostic significance in acute hepatocellular disorders.

The normal range for aminotransferases varies widely among laboratories, but generally ranges from 10 to 40 IU/L. The interlaboratory variation in normal range is due to technical reasons; no reference standards exist to establish upper limits of normal for ALT and AST. Some have recommended revisions of normal limits of the aminotransferases to adjust for sex and body mass index, but others have noted the potential costs and unclear benefits of implementing this change.

Any type of liver cell injury can cause modest elevations in the serum aminotransferases. Levels of up to 300 IU/L are nonspecific and may be found in any type of liver disorder. Minimal ALT elevations in asymptomatic blood donors rarely indicate severe liver disease; studies

have shown that fatty liver disease is the most likely explanation. Striking elevations—that is, aminotransferases >1000 IU/L—occur almost exclusively in disorders associated with extensive hepatocellular injury such as (1) viral hepatitis, (2) ischemic liver injury (prolonged hypotension or acute heart failure), or (3) toxin- or drug-induced liver injury.

The pattern of the aminotransferase elevation can be helpful diagnostically. In most acute hepatocellular disorders, the ALT is higher than or equal to the AST. Whereas the AST:ALT ratio is typically <1 in patients with chronic viral hepatitis and nonalcoholic fatty liver disease, a number of groups have noted that as cirrhosis develops, this ratio rises to >1. An AST:ALT ratio >2:1 is suggestive, whereas a ratio >3:1 is highly suggestive, of alcoholic liver disease. The AST in alcoholic liver disease is rarely >300 IU/L, and the ALT is often normal. A low level of ALT in the serum is due to an alcohol-induced deficiency of pyridoxal phosphate.

The aminotransferases are usually not greatly elevated in obstructive jaundice. One notable exception occurs during the acute phase of biliary obstruction caused by the passage of a gallstone into the common bile duct. In this setting, the aminotransferases can briefly be in the 1000–2000 IU/L range. However, aminotransferase levels decrease quickly, and the biochemical tests rapidly evolve into those typical of cholestasis.

ENZYMES THAT REFLECT CHOLESTASIS The activities of three enzymes—alkaline phosphatase, 5'-nucleotidase, and γ -glutamyl transpeptidase (GGT)—are usually elevated in cholestasis. Alkaline

phosphatase and 5'-nucleotidase are found in or near the bile canalicular membrane of hepatocytes, whereas GGT is located in the endoplasmic reticulum and in bile duct epithelial cells. Reflecting its more diffuse localization in the liver, GGT elevation in serum is less specific for cholestasis than are elevations of alkaline phosphatase or 5'-nucleotidase. Some have advocated the use of GGT to identify patients with occult alcohol use. Its lack of specificity makes its use in this setting questionable.

The normal serum alkaline phosphatase consists of many distinct isoenzymes found in the liver, bone, placenta, and, less commonly, the small intestine. Patients over age 60 can have a mildly elevated alkaline phosphatase (1–1.5 times normal), whereas individuals with blood types O and B can have an elevation of the serum alkaline phosphatase after eating a fatty meal due to the influx of intestinal alkaline phosphatase into the blood. It is also elevated in children and adolescents undergoing rapid bone growth because of bone alkaline phosphatase and late in normal pregnancies due to the influx of placental alkaline phosphatase.

Elevation of liver-derived alkaline phosphatase is not totally specific for cholestasis, and a less than threefold elevation can be seen in almost any type of liver disease. Alkaline phosphatase elevations greater than four times normal occur primarily in patients with cholestatic liver disorders, infiltrative liver diseases such as cancer and amyloidosis, and bone conditions characterized by rapid bone turnover (e.g., Paget's disease). In bone diseases, the elevation is due to increased amounts of the bone isoenzymes. In liver diseases, the elevation is almost always due to increased amounts of the liver isoenzyme.

If an elevated serum alkaline phosphatase is the only abnormal finding in an apparently healthy person or if the degree of elevation is higher than expected in the clinical setting, identification of the source of elevated isoenzymes is helpful (Fig. 330-1). This problem can be approached in two ways. First, and most precise, is the fractionation of the alkaline phosphatase by electrophoresis. The second, best substantiated, and most available approach involves the measurement of serum 5'-nucleotidase or GGT. These enzymes are rarely elevated in conditions other than liver disease.

In the absence of jaundice or elevated aminotransferases, an elevated alkaline phosphatase of liver origin often, but not always, suggests early cholestasis and, less often, hepatic infiltration by tumor or granulomata. Other conditions that cause isolated elevations of the alkaline phosphatase include primary biliary cholangitis, sclerosing cholangitis, Hodgkin's disease, diabetes, hyperthyroidism, congestive heart failure, and amyloidosis.

The level of serum alkaline phosphatase elevation is not helpful in distinguishing between intrahepatic and extrahepatic cholestasis. There is essentially no difference among the values found in obstructive jaundice due to cancer, common duct stone, sclerosing cholangitis, or bile duct stricture. Values are similarly increased in patients with intrahepatic cholestasis due to drug-induced hepatitis, primary biliary cholangitis, sepsis, rejection of transplanted livers, and, rarely, alcohol-induced steatohepatitis. Values are also greatly elevated in hepatobiliary disorders seen in patients with AIDS (e.g., AIDS cholangiopathy due to cytomegalovirus or cryptosporidial infection and tuberculosis with hepatic involvement).

■ TESTS THAT MEASURE BIOSYNTHETIC FUNCTION OF THE LIVER

Serum Albumin Serum albumin is synthesized exclusively by hepatocytes. Serum albumin has a long half-life: 18–20 days, with ~4% degraded per day. Because of this slow turnover, the serum albumin is not a good indicator of acute or mild hepatic dysfunction; only minimal changes in the serum albumin are seen in acute liver conditions such as viral hepatitis, drug-related hepatotoxicity, and obstructive jaundice. In hepatitis, albumin levels <3 g/dL should raise the possibility of chronic liver disease. Hypoalbuminemia is more common in chronic liver disorders such as cirrhosis and usually reflects severe liver damage and decreased albumin synthesis. However, hypoalbuminemia is not specific for liver disease and may occur in protein malnutrition of any

cause, as well as protein-losing enteropathies, nephrotic syndrome, and chronic infections that are associated with prolonged increases in levels of cytokines that inhibit albumin synthesis, such as serum interleukin 1 and/or tumor necrosis factor. Serum albumin should not be measured to screen patients in whom there is no suspicion of liver disease. A general medical clinic study of consecutive patients in whom no indications were present for albumin measurement showed that although 12% of patients had abnormal test results, the finding was of clinical importance in only 0.4%.

Serum Globulins Serum globulins are a group of proteins made up of γ globulins (immunoglobulins) produced by B lymphocytes and α and β globulins produced primarily in hepatocytes. γ Globulins are increased in chronic liver disease, such as chronic hepatitis and cirrhosis. In cirrhosis, the increased serum γ globulin concentration is due to the increased synthesis of antibodies, some of which are directed against intestinal bacteria. This occurs because the cirrhotic liver fails to clear bacterial antigens that normally reach the liver through the hepatic circulation.

Increases in the concentration of specific isotypes of γ globulins are often helpful in the recognition of certain chronic liver diseases. Diffuse polyclonal increases in IgG levels are common in autoimmune hepatitis; increases >100% should alert the clinician to this possibility. Increases in the IgM levels are common in primary biliary cholangitis, whereas increases in the IgA levels occur in alcoholic liver disease.

■ COAGULATION FACTORS

With the exception of factor VIII, which is produced by vascular endothelial cells, the blood clotting factors are made exclusively in hepatocytes. Their serum half-lives are much shorter than albumin, ranging from 6 h for factor VII to 5 days for fibrinogen. Because of their rapid turnover, measurement of the clotting factors is the single best acute measure of hepatic synthetic function and helpful in both diagnosis and assessing the prognosis of acute parenchymal liver disease. Useful for this purpose is the *serum prothrombin time*, which collectively measures factors II, V, VII, and X. Biosynthesis of factors II, VII, IX, and X depends on vitamin K. The international normalized ratio (INR) is used to express the degree of anticoagulation on warfarin therapy. The INR standardizes prothrombin time measurement according to the characteristics of the thromboplastin reagent used in a particular lab, which is expressed as an International Sensitivity Index (ISI); the ISI is then used in calculating the INR.

The prothrombin time may be elevated in hepatitis and cirrhosis as well as in disorders that lead to vitamin K deficiency such as obstructive jaundice or fat malabsorption of any kind. Marked prolongation of the prothrombin time, >5 s above control and not corrected by parenteral vitamin K administration, is a poor prognostic sign in acute viral hepatitis and other acute and chronic liver diseases. The INR, along with the total serum bilirubin and creatinine, are components of the MELD score, which is used as a measure of hepatic decompensation and to allocate organs for liver transplantation.

■ OTHER DIAGNOSTIC TESTS

Although tests may direct the physician to a category of liver disease, additional biochemical testing, radiologic testing, and procedures are often necessary to make the proper diagnosis, as shown in Fig. 337-1. The most commonly used ancillary tests are reviewed here, as are the noninvasive tests available for assessing hepatic fibrosis.

Ammonia Ammonia is produced in the body during normal protein metabolism and by intestinal bacteria, primarily those in the colon. The liver plays a role in the detoxification of ammonia by converting it to urea, which is excreted by the kidneys. Striated muscle also plays a role in detoxification of ammonia, where it is combined with glutamic acid to form glutamine. Patients with advanced liver disease typically have significant muscle wasting, which likely contributes to hyperammonemia. Some physicians use the blood ammonia for detecting encephalopathy or for monitoring hepatic synthetic function, although its use for either of these indications has problems. There is very poor correlation between either the presence or the severity

TABLE 337-1 Liver Test Patterns in Hepatobiliary Disorders

Type of Disorder	Bilirubin	Aminotransferases	Alkaline Phosphatase	Albumin	Prothrombin Time
Hemolysis/Gilbert's syndrome	Normal to 86 µmol/L (5 mg/dL) 85% due to indirect fractions No bilirubinuria	Normal	Normal	Normal	Normal
Acute hepatocellular necrosis (viral, ischemic, and drug- or toxin-induced hepatitis)	Both fractions may be elevated Peak usually follows aminotransferases Bilirubinuria	Elevated, often >500 IU, ALT > AST	Normal to <3x normal elevation	Normal	Usually normal. If >5x above control and not corrected by parenteral vitamin K, suggests poor prognosis
Chronic hepatocellular disorders	Both fractions may be elevated Bilirubinuria	Elevated, but usually <300 IU	Normal to <3x normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Alcoholic hepatitis, cirrhosis	Both fractions may be elevated Bilirubinuria	AST:ALT >2 suggests alcoholic hepatitis or cirrhosis	Normal to <3x normal elevation	Often decreased	Often prolonged Fails to correct with parenteral vitamin K
Intra- and extrahepatic cholestasis (obstructive jaundice)	Both fractions may be elevated Bilirubinuria	Normal to moderate elevation Rarely >500 IU	Elevated, often >4x normal elevation	Normal, unless chronic	Normal If prolonged, will correct with parenteral vitamin K
Infiltrative diseases (tumor, granulomatosis)	Usually normal	Normal to slight elevation	Elevated, often >4x normal elevation Fractionate, or confirm liver origin with 5'-nucleotidase or γ-glutamyl transpeptidase	Normal	Normal

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase.

of acute encephalopathy and elevation of blood ammonia; it can be occasionally useful for identifying occult liver disease in patients with mental status changes. There is also a poor correlation of the blood serum ammonia and hepatic function. The ammonia can be elevated in patients with severe portal hypertension and portal blood shunting around the liver even in the presence of normal or near-normal hepatic function. Elevated arterial ammonia levels have been shown to correlate with outcome in fulminant hepatic failure.

Liver Biopsy Percutaneous biopsy of the liver is a safe procedure that is easily performed with local anesthesia and ultrasound guidance. Liver biopsy is of proven value in the following situations: (1) hepatocellular disease of uncertain cause, (2) prolonged hepatitis with the possibility of autoimmune hepatitis, (3) unexplained hepatomegaly, (4) unexplained splenomegaly, (5) hepatic lesions uncharacterized by radiologic imaging, (6) fever of unknown origin, and (7) staging of malignant lymphoma. Liver biopsy is most accurate in disorders causing diffuse changes throughout the liver and is subject to sampling error in focal disorders. Liver biopsy should not be the initial procedure in the diagnosis of cholestasis. The biliary tree should first be assessed for signs of obstruction. Contraindications to performing a percutaneous liver biopsy include significant ascites and prolonged INR. Under these circumstances, the biopsy can be performed via the transjugular approach.

Noninvasive Tests to Detect Hepatic Fibrosis Although liver biopsy is the standard for the assessment of hepatic fibrosis, noninvasive measures of hepatic fibrosis have been developed and show promise. These measures include multiparameter tests aimed at detecting and staging the degree of hepatic fibrosis and imaging techniques. FibroTest (marketed as FibroSure in the United States) is the best evaluated of the multiparameter blood tests. The test incorporates haptoglobin, bilirubin, GGT, apolipoprotein A-I, and α₁-macroglobulin and has been found to have high positive and negative predictive values for diagnosing advanced fibrosis in patients with chronic hepatitis C, chronic hepatitis B, alcoholic liver disease, or nonalcoholic fatty liver disease and patients taking methotrexate for psoriasis. Transient elastography (TE), marketed as FibroScan, and magnetic resonance elastography (MRE) both have gained U.S. Food and Drug Administration approval for use in the management of patients with liver disease. TE uses ultrasound waves to measure hepatic stiffness noninvasively. TE

has been shown to be accurate for identifying advanced fibrosis in patients with chronic hepatitis C, primary biliary cholangitis, hemochromatosis, nonalcoholic fatty liver disease, and recurrent chronic hepatitis after liver transplantation. MRE has been found to be superior to TE for staging liver fibrosis in patients with a variety of chronic liver diseases but requires access to a magnetic resonance imaging scanner and is more expensive.

Ultrasonography Ultrasonography is the first diagnostic test to use in patients whose liver tests suggest cholestasis, to look for the presence of a dilated intrahepatic or extrahepatic biliary tree or to identify gallstones. In addition, it shows space-occupying lesions within the liver, enables the clinician to distinguish between cystic and solid masses, and helps direct percutaneous biopsies. Ultrasound with Doppler imaging can detect the patency of the portal vein, hepatic artery, and hepatic veins and determine the direction of blood flow. This is the first test ordered in patients suspected of having Budd-Chiari syndrome.

■ USE OF LIVER TESTS

As previously noted, the best way to increase the sensitivity and specificity of laboratory tests in the detection of liver disease is to employ a battery of tests that includes the aminotransferases, alkaline phosphatase, bilirubin, albumin, and prothrombin time along with the judicious use of the other tests described in this chapter. Table 337-1 shows how patterns of liver tests can lead the clinician to a category of disease that will direct further evaluation. However, it is important to remember that no single set of liver tests will necessarily provide a diagnosis. It is often necessary to repeat these tests on several occasions over days to weeks for a diagnostic pattern to emerge. Figure 337-1 is an algorithm for the evaluation of chronically abnormal liver tests.

■ GLOBAL CONSIDERATIONS

The tests and principles presented in this chapter are applicable worldwide. The causes of liver test abnormalities vary according to region. In developing nations, infectious diseases are more commonly the etiology of abnormal serum liver tests than in developed nations.

A

This chapter represents a revised version of a chapter in previous editions of *Harrison's* in which Marshall M. Kaplan was a co-author.

FURTHER READING

- K PS, K WR: The Model for End-Stage Liver Disease (MELD). *Hepatology* 45:797, 2007.
 K M: Alkaline phosphatase. *Gastroenterology* 62:452, 1972.
 M SM et al: Noninvasive assessment of liver fibrosis. *Hepatology* 53:325, 2011.
 P D et al: Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann Intern Med* 137:1, 2002.

338

The Hyperbilirubinemias

Allan W. Wolkoff

BILIRUBIN METABOLISM

The details of bilirubin metabolism are presented in [Chap. 49](#). However, the hyperbilirubinemias are best understood in terms of perturbations of specific aspects of bilirubin metabolism and transport, and these will be briefly reviewed here as depicted in [Fig. 338-1](#).

Bilirubin is the end product of heme degradation. Some 70–90% of bilirubin is derived from degradation of the hemoglobin of senescent red blood cells. Bilirubin produced in the periphery is transported to the liver within the plasma, where, due to its insolubility in aqueous solutions, it is tightly bound to albumin. Under normal circumstances, bilirubin is removed from the circulation rapidly and efficiently by hepatocytes. Transfer of bilirubin from blood to bile involves four distinct but interrelated steps (Fig. 338-1).

- Hepatocellular uptake:** Uptake of bilirubin by the hepatocyte has carrier-mediated kinetics. Although a number of candidate bilirubin transporters have been proposed, the identity of the actual transporter remains elusive.
- Intracellular binding:** Within the hepatocyte, bilirubin is kept in solution by binding as a nonsubstrate ligand to several of the glutathione-S-transferases, formerly called ligandins.

Conjugation: Bilirubin is conjugated with one or two glucuronic acid moieties by a specific UDP-glucuronosyltransferase to form bilirubin mono- and diglucuronide, respectively. Conjugation disrupts the internal hydrogen bonding that limits aqueous solubility of bilirubin, and the resulting glucuronide conjugates are highly soluble in water. Conjugation is obligatory for excretion of bilirubin across the bile canalicular membrane into bile. The UDP-glucuronosyltransferases have been classified into gene families based on the degree of homology among the mRNAs for the various isoforms. Those that conjugate bilirubin and certain other substrates have been designated the *UGT1* family. These are expressed from a single gene complex by alternative promoter usage. This gene complex contains multiple substrate-specific first exons, designated A1, A2, etc. ([Fig. 338-2](#)), each with its own promoter and each encoding the amino-terminal half of a specific isoform. In addition, there are four common exons (exons 2–5) that encode the shared carboxyl-terminal half of all of the *UGT1* isoforms. The various first exons encode the specific aglycone substrate binding sites for each isoform, while the shared exons encode the binding site for the sugar donor, UDP-glucuronic acid, and the transmembrane domain. Exon A1 and the four common exons, collectively designated as the *UGT1A1* gene ([Fig. 338-2](#)), encode the physiologically critical enzyme bilirubin-UDP-glucuronosyltransferase (*UGT1A1*). A functional corollary of the organization of the *UGT1* gene is that a mutation in one of the first exons will affect only a single enzyme isoform. By contrast, a mutation in exons 2–5 will alter all isoforms encoded by the *UGT1* gene complex.

Biliary excretion: It has been thought until recently that bilirubin mono- and diglucuronides are excreted directly across the canalicular plasma membrane into the bile canaliculus by an ATP-dependent transport process mediated by a canalicular membrane protein called *multidrug resistance-associated protein 2* (MRP2, ABCC2). Mutations of MRP2 result in the Dubin-Johnson syndrome (see below). However, studies in patients with Rotor syndrome (see below) indicate that after formation, a portion of the glucuronides is transported into the portal circulation by a sinusoidal membrane protein called *multidrug resistance-associated protein 3* (MRP3, ABCC3) and is subjected to reuptake into the hepatocyte by the sinusoidal membrane uptake transporters *organic anion transport protein 1B1* (OATP1B1, SLC01B1) and OATP1B3 (SLCO1B3).

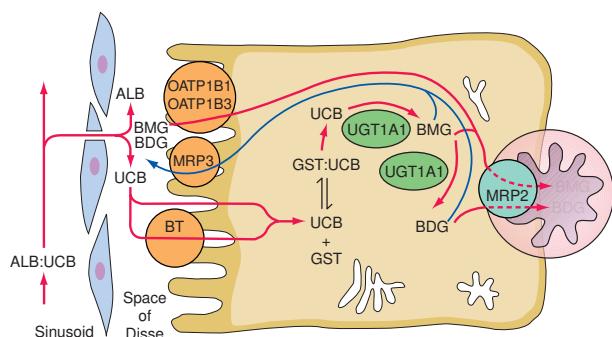


FIGURE 338-1 Hepatocellular bilirubin transport. Albumin-bound bilirubin in sinusoidal blood passes through endothelial cell fenestrae to reach the hepatocyte surface, entering the cell by both facilitated and simple diffusional processes. Within the cell, it is bound to glutathione-S-transferases and conjugated by bilirubin-UDP-glucuronosyltransferase (UGT1A1) to mono- and diglucuronides, which are actively transported across the canalicular membrane into the bile. In addition to this direct excretion of bilirubin glucuronides, a portion are transported into the portal circulation by MRP3 and subjected to reuptake into the hepatocyte by OATP1B1 and OATP1B3. ALB, albumin; BDG, bilirubin diglucuronide; BMG, bilirubin monoglucuronide; BT, proposed bilirubin transporter; GST, glutathione-S-transferase; MRP2 and MRP3, multidrug resistance-associated proteins 2 and 3; OATP1B1 and OATP1B3, organic anion transport proteins 1B1 and 1B3; UCB, unconjugated bilirubin; UGT1A1, bilirubin-UDP-glucuronosyltransferase.

EXTRAHEPATIC ASPECTS OF BILIRUBIN DISPOSITION

Bilirubin in the Gut Following secretion into bile, conjugated bilirubin reaches the duodenum and passes down the gastrointestinal tract without reabsorption by the intestinal mucosa. An appreciable fraction is converted by bacterial metabolism in the gut to the water-soluble colorless compound urobilinogen. Urobilinogen undergoes enterohepatic cycling. Urobilinogen not taken up by the liver reaches the systemic circulation, from which some is cleared by the kidneys. Unconjugated bilirubin ordinarily does not reach the gut except in neonates or, by ill-defined alternative pathways, in the presence of severe unconjugated hyperbilirubinemia (e.g., Crigler-Najjar syndrome, type I [$CN-I$]). Unconjugated bilirubin that reaches the gut is partly reabsorbed, amplifying any underlying hyperbilirubinemia.

Renal Excretion of Bilirubin Conjugates Unconjugated bilirubin is not excreted in urine, as it is too tightly bound to albumin for effective glomerular filtration and there is no tubular mechanism for its renal secretion. In contrast, the bilirubin conjugates are readily filtered at the glomerulus and can appear in urine in disorders characterized by increased bilirubin conjugates in the circulation. It should be kept in mind that the kidney can serve as an ‘overflow valve’ for conjugated bilirubin. Consequently, the level of jaundice in individuals with conjugated hyperbilirubinemia can be amplified in the presence of renal failure.

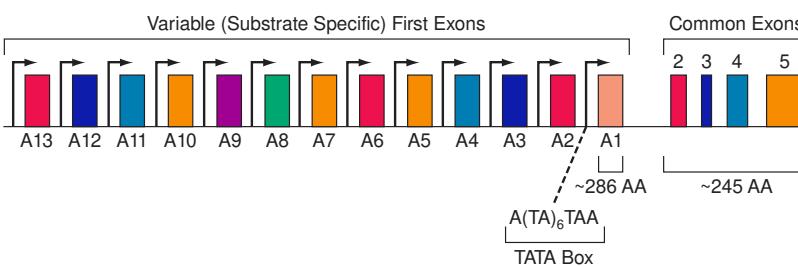


FIGURE 338-2 Structural organization of the human *UGT1* gene complex. This large complex on chromosome 2 contains at least 13 substrate-specific first exons (A1, A2, etc.). Since four of these are pseudogenes, nine *UGT1* isoforms with differing substrate specificities are expressed. Each exon 1 has its own promoter and encodes the amino-terminal substrate-specific 286 amino acids of the various *UGT1*-encoded isoforms, and common exons 2–5 encode the 245 carboxyl-terminal amino acids common to all of the isoforms. mRNAs for specific isoforms are assembled by splicing a particular first exon such as the bilirubin-specific exon A1 to exons 2 to 5. The resulting message encodes a complete enzyme, in this particular case, bilirubin-UDP-glucuronosyltransferase (*UGT1A1*). Mutations in a first exon affect only a single isoform. Those in exons 2–5 affect all enzymes encoded by the *UGT1* complex.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO UNCONJUGATED HYPERBILIRUBINEMIA

■ INCREASED BILIRUBIN PRODUCTION

Hemolysis Increased destruction of erythrocytes leads to increased bilirubin turnover and unconjugated hyperbilirubinemia; the hyperbilirubinemia is usually modest in the presence of normal liver function. In particular, the bone marrow is only capable of a sustained eightfold increase in erythrocyte production in response to a hemolytic stress. Therefore, hemolysis alone cannot result in a sustained hyperbilirubinemia of more than 68 μmol/L (4 mg/dL). Higher values imply concomitant hepatic dysfunction. When hemolysis is the only abnormality in an otherwise healthy individual, the result is a purely unconjugated hyperbilirubinemia, with the direct-reacting fraction as measured in a typical clinical laboratory being ≤15% of the total serum bilirubin. In the presence of systemic disease, which may include a degree of hepatic dysfunction, hemolysis may produce a component of conjugated hyperbilirubinemia in addition to an elevated unconjugated bilirubin concentration. Prolonged hemolysis may lead to the precipitation of bilirubin salts within the gallbladder or biliary tree, resulting in the formation of gallstones in which bilirubin, rather than cholesterol, is the major component. Such pigment stones may lead to acute or chronic cholecystitis, biliary obstruction, or any other biliary tract consequence of calculous disease.

Ineffective Erythropoiesis During erythroid maturation, small amounts of hemoglobin may be lost at the time of nuclear extrusion, and a fraction of developing erythroid cells is destroyed within the marrow. These processes normally account for a small proportion of bilirubin that is produced. In various disorders, including thalassemia major, megaloblastic anemias due to folate or vitamin B₁₂ deficiency, congenital erythropoietic porphyria, lead poisoning, and various congenital and acquired dyserythropoietic anemias, the fraction of total bilirubin production derived from ineffective erythropoiesis is increased, reaching as much as 70% of the total. This may be sufficient to produce modest degrees of unconjugated hyperbilirubinemia.

Miscellaneous Degradation of the hemoglobin of extravascular collections of erythrocytes, such as those seen in massive tissue infarctions or large hematomas, may lead transiently to unconjugated hyperbilirubinemia.

■ DECREASED HEPATIC BILIRUBIN CLEARANCE

Decreased Hepatic Uptake Decreased hepatic bilirubin uptake is believed to contribute to the unconjugated hyperbilirubinemia of Gilbert's syndrome (GS), although the molecular basis for this finding remains unclear (see below). Several drugs, including flavaspidic acid,

novobiocin, and rifampin, as well as various cholecystographic contrast agents, have been reported to inhibit bilirubin uptake. The resulting unconjugated hyperbilirubinemia resolves with cessation of the medication.

Impaired Conjugation • PHYSIOLOGIC NEONATAL JAUNDICE Bilirubin produced by the fetus is cleared by the placenta and eliminated by the maternal liver. Immediately after birth, the neonatal liver must assume responsibility for bilirubin clearance and excretion. However, many hepatic physiologic processes are incompletely developed at birth. Levels of *UGT1A1* are low, and alternative excretory pathways allow passage of unconjugated bilirubin into the gut. Since the intestinal flora that convert bilirubin to urobilinogen are also

undeveloped, an enterohepatic circulation of unconjugated bilirubin ensues. As a consequence, most neonates develop mild unconjugated hyperbilirubinemia between days 2 and 5 after birth. Peak levels are typically <85–170 μmol/L (5–10 mg/dL) and decline to normal adult concentrations within 2 weeks, as mechanisms required for bilirubin disposition mature. Prematurity, often associated with more profound immaturity of hepatic function and hemolysis, can result in higher levels of unconjugated hyperbilirubinemia. A rapidly rising unconjugated bilirubin concentration, or absolute levels >340 μmol/L (20 mg/dL), puts the infant at risk for bilirubin encephalopathy, or kernicterus. Under these circumstances, bilirubin crosses an immature blood-brain barrier and precipitates in the basal ganglia and other areas of the brain. The consequences range from appreciable neurologic deficits to death. Treatment options include phototherapy, which converts bilirubin into water-soluble photoisomers that are excreted directly into bile, and exchange transfusion. The canalicular mechanisms responsible for bilirubin excretion are also immature at birth, and their maturation may lag behind that of *UGT1A1*; this can lead to transient conjugated neonatal hyperbilirubinemia, especially in infants with hemolysis.

ACQUIRED CONJUGATION DEFECTS A modest reduction in bilirubin conjugating capacity may be observed in advanced hepatitis or cirrhosis. However, in this setting, conjugation is better preserved than other aspects of bilirubin disposition, such as canalicular excretion. Various drugs, including pregnanediol, novobiocin, chloramphenicol, gentamicin, and atazanavir, may produce unconjugated hyperbilirubinemia by inhibiting *UGT1A1* activity. Bilirubin conjugation may be inhibited by certain fatty acids that are present in breast milk, but not serum, of mothers whose infants have excessive neonatal hyperbilirubinemia (*breast milk jaundice*). Alternatively, there may be increased enterohepatic circulation of bilirubin in these infants. The pathogenesis of breast milk jaundice appears to differ from that of transient familial neonatal hyperbilirubinemia (Lucey-Driscoll syndrome), in which there may be a *UGT1A1* inhibitor in maternal serum.

■ HEREDITARY DEFECTS IN BILIRUBIN CONJUGATION

Three familial disorders characterized by differing degrees of unconjugated hyperbilirubinemia have long been recognized. The defining clinical features of each are described below (Table 338-1). While these disorders have been recognized for decades to reflect differing degrees of deficiency in the ability to conjugate bilirubin, recent advances in the molecular biology of the *UGT1* gene complex have elucidated their interrelationships and clarified previously puzzling features.

Crigler-Najjar Syndrome, Type I CN-I is characterized by striking unconjugated hyperbilirubinemia of 340–765 μmol/L (20–45 mg/dL) that appears in the neonatal period and persists for life. Other conventional hepatic biochemical tests such as serum aminotransferases

TABLE 338-1 Principal Differential Characteristics of Gilbert and Crigler-Najjar Syndromes

FEATURE	CRIGLER-NAJJAR SYNDROME		GILBERT SYNDROME
	TYPE I	TYPE II	
Total serum bilirubin, $\mu\text{mol/L}$ (mg/dL)	310–755 (usually >345) (18–45 [usually >20])	100–430 (usually ≤345) (6–25 [usually ≤20])	Typically ≤70 $\mu\text{mol/L}$ (<4 mg/dL) in absence of fasting or hemolysis
Routine liver tests	Normal	Normal	Normal
Response to phenobarbital	None	Decreases bilirubin by >25%	Decreases bilirubin to normal
Kernicterus	Usual	Rare	No
Hepatic histology	Normal	Normal	Usually normal; increased lipofuscin pigment in some
Bile characteristics			
Color	Pale or colorless	Pigmented	Normal dark color
Bilirubin fractions	>90% unconjugated	Largest fraction (mean: 57%) monoconjugates	Mainly diconjugates but monoconjugates increased (mean: 23%)
Bilirubin UDP-glucuronosyltransferase activity	Typically absent; traces in some patients	Markedly reduced: 0–10% of normal	Reduced: typically 10–33% of normal Promoter mutation: recessive
Inheritance (all autosomal)	Recessive	Predominantly recessive	Missense mutations: 7 of 8 dominant; 1 reportedly recessive

and alkaline phosphatase are normal, and there is no evidence of hemolysis. Hepatic histology is also essentially normal except for the occasional presence of bile plugs within canaliculi. Bilirubin glucuronides are virtually absent from the bile, and there is no detectable constitutive expression of UGT1A1 activity in hepatic tissue. Neither UGT1A1 activity nor the serum bilirubin concentration responds to administration of phenobarbital or other enzyme inducers. Unconjugated bilirubin accumulates in plasma, from which it is eliminated very slowly by alternative pathways that include direct passage into the bile and small intestine, possibly via bilirubin photoisomers. This accounts for the small amount of urobilinogen found in feces. No bilirubin is found in the urine. First described in 1952, the disorder is rare (estimated prevalence, 0.6–1.0 per million). Many patients are from geographically or socially isolated communities in which consanguinity is common, and pedigree analyses show an autosomal recessive pattern of inheritance. The majority of patients (type IA) exhibit defects in the glucuronide conjugation of a spectrum of substrates in addition to bilirubin, including various drugs and other xenobiotics. These individuals have mutations in one of the common exons (2–5) of the *UGT1* gene (Fig. 338-2). In a smaller subset (type IB), the defect is limited largely to bilirubin conjugation, and the causative mutation is in the bilirubin-specific exon A1. Estrogen glucuronidation is mediated by UGT1A1 and is defective in all CN-I patients. More than 30 different genetic lesions of *UGT1A1* responsible for CN-I have been identified, including deletions, insertions, alterations in intron splice donor and acceptor sites, exon skipping, and point mutations that introduce premature stop codons or alter critical amino acids. Their common feature is that they all encode proteins with absent or, at most, traces of bilirubin-UDP-glucuronosyltransferase enzymatic activity.

Prior to the use of phototherapy, most patients with CN-I died of bilirubin encephalopathy (*kernicterus*) in infancy or early childhood. A few lived as long as early adult life without overt neurologic damage, although more subtle testing usually indicated mild but progressive brain damage. In the absence of liver transplantation, death eventually supervened from late-onset bilirubin encephalopathy, which often followed a nonspecific febrile illness. Although isolated hepatocyte transplantation has been used in a small number of cases of CN-I, early liver transplantation (Chap. 345) remains the best hope to prevent brain injury and death at present. It is anticipated that gene replacement therapy may be an option in the future.

Crigler-Najjar Syndrome, Type II (CN-II) This condition was recognized as a distinct entity in 1962 and is characterized by marked unconjugated hyperbilirubinemia in the absence of abnormalities of other conventional hepatic biochemical tests, hepatic histology, or hemolysis. It differs from CN-I in several specific ways (Table 338-1): (1) although there is considerable overlap, average

bilirubin concentrations are lower in CN-II; (2) accordingly, CN-II is only infrequently associated with kernicterus; (3) bile is deeply colored, and bilirubin glucuronides are present, with a striking, characteristic increase in the proportion of monoglucuronides; (4) UGT1A1 in liver is usually present at reduced levels (typically ≤10% of normal); and (5) while typically detected in infancy, hyperbilirubinemia was not recognized in some cases until later in life and, in one instance, at age 34. As with CN-I, most CN-II cases exhibit abnormalities in the conjugation of other compounds, such as salicylamide and menthol, but in some instances, the defect appears limited to bilirubin. Reduction of serum bilirubin concentrations by >25% in response to enzyme inducers such as phenobarbital distinguishes CN-II from CN-I, although this response may not be elicited in early infancy and often is not accompanied by measurable UGT1A1 induction. Bilirubin concentrations during phenobarbital administration do not return to normal but are typically in the range of 51–86 $\mu\text{mol/L}$ (3–5 mg/dL). Although the incidence of kernicterus in CN-II is low, instances have occurred, not only in infants but also in adolescents and adults, often in the setting of an intercurrent illness, fasting, or another factor that temporarily raises the serum bilirubin concentration above baseline and reduces serum albumin levels. For this reason, phenobarbital therapy is widely recommended, a single bedtime dose often sufficing to maintain clinically safe serum bilirubin concentrations.

Over 100 different mutations in the *UGT1* gene have been identified as causing CN-I or CN-II. It was found that missense mutations are more common in CN-II patients, as would be expected in this less severe phenotype. Their common feature is that they encode for a bilirubin-UDP-glucuronosyltransferase with markedly reduced, but detectable, enzymatic activity. The spectrum of residual enzyme activity explains the spectrum of phenotypic severity of the resulting hyperbilirubinemia. Molecular analysis has established that a large majority of CN-II patients are either homozygotes or compound heterozygotes for CN-II mutations and that individuals carrying one mutated and one entirely normal allele have normal bilirubin concentrations.

Gilbert Syndrome This syndrome is characterized by mild unconjugated hyperbilirubinemia, normal values for standard hepatic biochemical tests, and normal hepatic histology other than a modest increase of lipofuscin pigment in some patients. Serum bilirubin concentrations are most often <51 $\mu\text{mol/L}$ (<3 mg/dL), although both higher and lower values are frequent. The clinical spectrum of hyperbilirubinemia fades into that of CN-II at serum bilirubin concentrations of 86–136 $\mu\text{mol/L}$ (5–8 mg/dL). At the other end of the scale, the distinction between mild cases of GS and a normal state is often blurred. Bilirubin concentrations may fluctuate substantially in any given individual, and at least 25% of patients will exhibit temporarily normal values during prolonged follow-up. More elevated values are

associated with stress, fatigue, alcohol use, reduced caloric intake, and intercurrent illness, while increased caloric intake or administration of enzyme-inducing agents produces lower bilirubin levels. GS is most often diagnosed at or shortly after puberty or in adult life during routine examinations that include multichannel biochemical analyses. UGT1A1 activity is typically reduced to 10–35% of normal, and bile pigments exhibit a characteristic increase in bilirubin monoglycuronides. Studies of radiobilirubin kinetics indicate that hepatic bilirubin clearance is reduced to an average of one-third of normal. Administration of phenobarbital normalizes both the serum bilirubin concentration and hepatic bilirubin clearance; however, failure of UGT1A1 activity to improve in many such instances suggests the possible coexistence of an additional defect. Compartmental analysis of bilirubin kinetic data suggests that GS patients may have a defect in bilirubin uptake as well as in conjugation, although this has not been shown directly. Defects in the hepatic uptake of other organic anions that at least partially share an uptake mechanism with bilirubin, such as sulfobromophthalein and indocyanine green (ICG), are observed in a minority of patients. The metabolism and transport of bile acids that do not utilize the bilirubin uptake mechanism are normal. The magnitude of changes in the serum bilirubin concentration induced by provocation tests such as 48 h of fasting or the IV administration of nicotinic acid has been reported to be of help in separating GS patients from normal individuals. Other studies dispute this assertion. Moreover, on theoretical grounds, the results of such studies should provide no more information than simple measurements of the baseline serum bilirubin concentration. Family studies indicate that GS and hereditary hemolytic anemias such as hereditary spherocytosis, glucose-6-phosphate dehydrogenase deficiency, and β -thalassemia trait sort independently. Reports of hemolysis in up to 50% of GS patients are believed to reflect better case finding, since patients with both GS and hemolysis have higher bilirubin concentrations and are more likely to be jaundiced than patients with either defect alone.

GS is common, with many series placing its prevalence as high as 8%. Males predominate over females by reported ratios ranging from 1.5:1 to >7:1. However, these ratios may have a large artifactual component since normal males have higher mean bilirubin levels than normal females, but the diagnosis of GS is often based on comparison to normal ranges established in men. The high prevalence of GS in the general population may explain the reported frequency of mild unconjugated hyperbilirubinemia in liver transplant recipients. The disposition of most xenobiotics metabolized by glucuronidation appears to be normal in GS, as is oxidative drug metabolism in the majority of reported studies. The principal exception is the metabolism of the anti-tumor agent irinotecan (CPT-11), whose active metabolite (SN-38) is glucuronidated specifically by bilirubin-UDP-glucuronosyltransferase. Administration of CPT-11 to patients with GS has resulted in several toxicities, including intractable diarrhea and myelosuppression. Some reports also suggest abnormal disposition of menthol, estradiol benzoate, acetaminophen, tolbutamide, and rifamycin SV. Although some of these studies have been disputed, and there have been no reports of clinical complications from use of these agents in GS, prudence should be exercised in prescribing them or any agents metabolized primarily by glucuronidation in this condition. It should also be noted that the HIV protease inhibitors indinavir and atazanavir (*Chap. 202*) can inhibit UGT1A1, resulting in hyperbilirubinemia that is most pronounced in patients with preexisting GS.

Most older pedigree studies of GS were consistent with autosomal dominant inheritance with variable expressivity. However, studies of the *UGT1* gene in GS have indicated a variety of molecular genetic bases for the phenotypic picture and several different patterns of inheritance. Studies in Europe and the United States found that nearly all patients had normal coding regions for UGT1A1 but were homozygous for the insertion of an extra TA (i.e., A[TA]₇TAA rather than A[TA]₆TAA) in the promoter region of the first exon. This appeared to be necessary, but not sufficient, for clinically expressed GS, since 15% of normal controls were also homozygous for this variant. While normal by standard criteria, these individuals had somewhat higher bilirubin concentrations than the rest of the controls studied. Heterozygotes for this abnormality had bilirubin concentrations identical

to those homozygous for the normal A[TA]₆TAA allele. The prevalence of the A[TA]₇TAA allele in a general Western population is 30%, in which case 9% would be homozygotes. This is slightly higher than the prevalence of GS based on purely phenotypic parameters. It was suggested that additional variables, such as mild hemolysis or a defect in bilirubin uptake, might be among the factors enhancing phenotypic expression of the defect.

Phenotypic expression of GS due solely to the A[TA]₇TAA promoter abnormality is inherited as an autosomal recessive trait. A number of CN-II kindreds have been identified in whom there is also an allele containing a normal coding region but the A[TA]₇TAA promoter abnormality. CN-II heterozygotes, who have the A[TA]₆TAA promoter, are phenotypically normal, whereas those with the A[TA]₇TAA promoter express the phenotypic picture of GS. GS in such kindreds may also result from homozygosity for the A[TA]₇TAA promoter abnormality. Seven different missense mutations in the *UGT1* gene that reportedly cause GS with dominant inheritance have been found in Japanese individuals. Another Japanese patient with mild unconjugated hyperbilirubinemia was homozygous for a missense mutation in exon 5. GS in her family appeared to be recessive.

DISORDERS OF BILIRUBIN METABOLISM LEADING TO MIXED OR PREDOMINANTLY CONJUGATED HYPERBILIRUBINEMIA

In hyperbilirubinemia due to acquired liver disease (e.g., acute hepatitis, common bile duct stone), there are usually elevations in the serum concentrations of both conjugated and unconjugated bilirubin. Although biliary tract obstruction or hepatocellular cholestatic injury may present on occasion with a predominantly conjugated hyperbilirubinemia, it is generally not possible to differentiate intrahepatic from extrahepatic causes of jaundice based on the serum levels or relative proportions of unconjugated and conjugated bilirubin. The major reason for determining the amounts of conjugated and unconjugated bilirubin in the serum is for the initial differentiation of hepatic parenchymal and obstructive disorders (mixed conjugated and unconjugated hyperbilirubinemia) from the inheritable and hemolytic disorders discussed above that are associated with unconjugated hyperbilirubinemia.

FAMILIAL DEFECTS IN HEPATIC EXCRETORY FUNCTION

Dubin-Johnson Syndrome (DJS) This benign, relatively rare disorder is characterized by low-grade, predominantly conjugated hyperbilirubinemia (*Table 338-2*). Total bilirubin concentrations are typically between 34 and 85 $\mu\text{mol/L}$ (2 and 5 mg/dL) but on occasion can be in the normal range or as high as 340–430 $\mu\text{mol/L}$ (20–25 mg/dL) and can fluctuate widely in any given patient. The degree of hyperbilirubinemia may be increased by intercurrent illness, oral contraceptive use, and pregnancy. Because the hyperbilirubinemia is due to a predominant rise in conjugated bilirubin, bilirubinuria is characteristically present. Aside from elevated serum bilirubin levels, other routine laboratory tests are normal. Physical examination is usually normal except for jaundice, although an occasional patient may have hepatosplenomegaly.

Patients with DJS are usually asymptomatic, although some may have vague constitutional symptoms. These latter patients have usually undergone extensive diagnostic examinations for unexplained jaundice and have high levels of anxiety. In women, the condition may be subclinical until the patient becomes pregnant or receives oral contraceptives, at which time chemical hyperbilirubinemia becomes frank jaundice. Even in these situations, other routine liver function tests, including serum alkaline phosphatase and transaminase activities, are normal.

A cardinal feature of DJS is the accumulation of dark, coarsely granular pigment in the lysosomes of centrilobular hepatocytes. As a result, the liver may be grossly black in appearance. This pigment is thought to be derived from epinephrine metabolites that are not excreted normally. The pigment may disappear during bouts of viral hepatitis, only to reaccumulate slowly after recovery.

TABLE 338-2 Principal Differential Characteristics of Inheritable Disorders of Bile Canicular Function

	DJS	ROTOR	PFIC1	BRIC1	PFIC2	BRIC2	PFIC3
Gene	<i>ABCCA</i>	<i>SLCO1B1/SLCO1B3</i>	<i>ATP8B1</i>	<i>ATP8B1</i>	<i>ABCB11</i>	<i>ABCB11</i>	<i>ABCB4</i>
Protein	MRP2	OATP1B1/1B3	FIC1	FIC1	BSEP	BSEP	MDR3
Cholestasis	No	No	Yes	Episodic	Yes	Episodic	Yes
Serum GGT	Normal	Normal	Normal	Normal	Normal	Normal	↑↑
Serum bile acids	Normal	Normal	↑↑	↑↑ during episodes	↑↑	↑↑ during episodes	↑↑
Clinical features	Mild conjugated hyperbilirubinemia; otherwise, normal liver function; dark pigment in liver; characteristic pattern of urinary coproporphyrins	Mild conjugated hyperbilirubinemia; otherwise, normal liver function; liver without abnormal pigmentation	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood	Recurrent episodes of cholestasis beginning at any age	Severe cholestasis beginning in childhood; decreased phospholipids in bile

Abbreviations: BRIC, benign recurrent intrahepatic cholestasis; BSEP, bile salt excretory protein; DJS, Dubin-Johnson syndrome; GGT, γ-glutamyl transferase; MRP2, multidrug resistance-associated protein 2; OATP1A/1B, organic anion transport proteins 1B1 and 1B3; PFIC, progressive familial intrahepatic cholestasis; ↑↑, increased.

Biliary excretion of a number of anionic compounds is compromised in DJS. These include various cholecystographic agents, as well as sulfobromophthalein (Bromsulphalein [BSP]), a synthetic dye formerly used in a test of liver function. In this test, the rate of disappearance of BSP from plasma was determined following bolus IV administration. BSP is conjugated with glutathione in the hepatocyte; the resulting conjugate is normally excreted rapidly into the bile canaliculus. Patients with DJS exhibit characteristic rises in plasma concentrations at 90 min after injection, due to reflux of conjugated BSP into the circulation from the hepatocyte. Dyes such as ICG that are taken up by hepatocytes but are not further metabolized prior to biliary excretion do not show this reflux phenomenon. Continuous BSP infusion studies suggest a reduction in the time to maximum plasma concentration (t_{max}) for biliary excretion. Bile acid disposition, including hepatocellular uptake and biliary excretion, is normal in DJS. These patients have normal serum and biliary bile acid concentrations and do not have pruritis.

By analogy with findings in several mutant rat strains, the selective defect in biliary excretion of bilirubin conjugates and certain other classes of organic compounds, but not of bile acids, that characterizes DJS in humans was found to reflect defective expression of MRP2 (ABCC2), an ATP-dependent canalicular membrane transporter. Several different mutations in the *ABCC2* gene produce the Dubin-Johnson phenotype, which has an autosomal recessive pattern of inheritance. Although MRP2 is undoubtedly important in the biliary excretion of conjugated bilirubin, the fact that this pigment is still excreted in the absence of MRP2 suggests that other, as yet uncharacterized, transport proteins may serve in a secondary role in this process.

Patients with DJS also have a diagnostic abnormality in urinary coproporphyrin excretion. There are two naturally occurring coproporphyrin isomers, I and III. Normally, 75% of the coproporphyrin in urine is isomer III. In urine from DJS patients, total coproporphyrin content is normal, but >80% is isomer I. Heterozygotes for the syndrome show an intermediate pattern. The molecular basis for this phenomenon remains unclear.

Rotor Syndrome (RS) This benign, autosomal recessive disorder is clinically similar to DJS (Table 338-2), although it is seen even less frequently. A major phenotypic difference is that the liver in patients with RS has no increased pigmentation and appears totally normal. The only abnormality in routine laboratory tests is an elevation of total serum bilirubin, due to a predominant rise in conjugated bilirubin. This is accompanied by bilirubinuria. Several additional features differentiate RS from DJS. In RS, the gallbladder is usually visualized on oral cholecystography, in contrast to the nonvisualization that is typical of DJS. The pattern of urinary coproporphyrin excretion also differs. The pattern in RS resembles that of many acquired disorders of hepatobiliary function, in which coproporphyrin I, the major coproporphyrin isomer in bile, refluxes from the hepatocyte back into the circulation and is excreted in urine. Thus, total urinary coproporphyrin excretion is substantially increased in RS, in contrast to the normal levels seen in

DJS. Although the fraction of coproporphyrin I in urine is elevated, it is usually <70% of the total, compared with ≥80% in DJS. The disorders also can be distinguished by their patterns of BSP excretion. Although clearance of BSP from plasma is delayed in RS, there is no reflux of conjugated BSP back into the circulation as seen in DJS. Kinetic analysis of plasma BSP infusion studies suggests the presence of a defect in intrahepatocellular storage of this compound. This has never been demonstrated directly. Recent studies indicate that the molecular basis of RS results from simultaneous deficiency of the hepatocyte plasma membrane transporters OATP1B1 (SLCO1B1) and OATP1B3 (SLCO1B3). This results in reduced reuptake by these transporters of conjugated bilirubin that has been pumped out of the hepatocyte into the portal circulation by MRP3 (ABCC3) (Fig. 338-1).

Benign Recurrent Intrahepatic Cholestasis (BRIC) This rare disorder is characterized by recurrent attacks of pruritus and jaundice. The typical episode begins with mild malaise and elevations in serum aminotransferase levels, followed rapidly by rises in alkaline phosphatase and conjugated bilirubin and onset of jaundice and itching. The first one or two episodes may be misdiagnosed as acute viral hepatitis. The cholestatic episodes, which may begin in childhood or adulthood, can vary in duration from several weeks to months, followed by a complete clinical and biochemical resolution. Intervals between attacks may vary from several months to years. Between episodes, physical examination is normal, as are serum levels of bile acids, bilirubin, transaminases, and alkaline phosphatase. The disorder is familial and has an autosomal recessive pattern of inheritance. BRIC is considered a benign disorder in that it does not lead to cirrhosis or end-stage liver disease. However, the episodes of jaundice and pruritus can be prolonged and debilitating, and some patients have undergone liver transplantation to relieve the intractable and disabling symptoms. Treatment during the cholestatic episodes is symptomatic; there is no specific treatment to prevent or shorten the occurrence of episodes.

A gene termed *FIC1* was recently identified and found to be mutated in patients with BRIC. Curiously, this gene is expressed strongly in the small intestine but only weakly in the liver. The protein encoded by *FIC1* shows little similarity to those that have been shown to play a role in bile canalicular excretion of various compounds. Rather, it appears to be a member of a P-type ATPase family that transports aminophospholipids from the outer to the inner leaflet of a variety of cell membranes. Its relationship to the pathobiology of this disorder remains unclear. A second phenotypically identical form of BRIC, termed BRIC type 2, has been described resulting from mutations in the bile salt excretory protein (BSEP), the protein that is defective in progressive familial intrahepatic cholestasis (PFIC) type 2 (Table 338-2). How some mutations in this protein result in the episodic BRIC phenotype is unknown.

Progressive Familial Intrahepatic Cholestasis This name is applied to three phenotypically related syndromes (Table 338-2). PFIC type 1 (Byler's disease) presents in early infancy as cholestasis that

may be initially episodic. However, in contrast to BRIC, Byler's disease progresses to malnutrition, growth retardation, and end-stage liver disease during childhood. This disorder is also a consequence of an *FIC1* mutation. The functional relationship of the FIC1 protein to the pathogenesis of cholestasis in these disorders is unknown. Two other types of PFIC (types 2 and 3) have been described. PFIC type 2 is associated with a mutation in the protein originally named *sister of P-glycoprotein*, now known as *bile salt excretory protein* (BSEP, ABCB11), which is the major bile canalicular exporter of bile acids. As noted above, some mutations of this protein are associated with BRIC type 2, rather than the PFIC type 2 phenotype. PFIC type 3 has been associated with a mutation of MDR3 (ABCB4), a protein that is essential for normal hepatocellular excretion of phospholipids across the bile canalculus. Although all three types of PFIC have similar clinical phenotypes, only type 3 is associated with high serum levels of γ -glutamyl transferase (GGT) activity. In contrast, activity of this enzyme is normal or only mildly elevated in symptomatic BRIC and PFIC types 1 and 2. Interestingly, mutations in *FIC1* or *BSEP* are not found in approximately one-third of patients with clinical PFIC and normal GGT. Recent studies have shown that patients with mutations in *NR1H4*, the gene encoding the farnesoid X receptor (FXR), a nuclear hormone receptor activated by bile acids, have a syndrome identical to PFIC2 with absent expression of BSEP. Mutations in tight junction protein 2 (TJP2) have also been associated with severe cholestasis with normal GGT levels, likely due to disruption of tight junctions at the bile canalculus.

FURTHER READING

- B LN, T RJ: Progressive familial intrahepatic cholestasis. *Clin Liver Dis* 22:657, 2018.
- C G et al: Gilbert and Crigler Najjar syndromes: An update of the UDP-glucuronosyltransferase 1A1 (UGT1A1) gene mutation database. *Blood Cells Mol Dis* 50:273, 2013.
- G -O N et al: Mutations in the nuclear bile acid receptor FXR cause progressive familial intrahepatic cholestasis. *Nat Commun* 7:10713, 2016.
- H TW: Biology of bilirubin photoisomers. *Clin Perinatol* 43:277, 2016.
- L AA: A pharmacologic view of phototherapy. *Clin Perinatol* 43:259, 2016.
- M N et al: Inherited disorders of bilirubin clearance. *Pediatr Res* 79:378, 2016.
- S M et al: Mutations in TJP2 cause progressive cholestatic liver disease. *Nat Genet* 46:326, 2014.
- S CJ, B JL: Biosynthesis and trafficking of the bile salt export pump, BSEP: Therapeutic implications of BSEP mutations. *Mol Aspects Med* 37:3, 2014.
- S E et al: Complete OATP1B1 and OATP1B3 deficiency causes human Rotor syndrome by interrupting conjugated bilirubin reuptake into the liver. *J Clin Invest* 122:519, 2012.
- W DBE et al: Genotype correlates with the natural history of severe bile salt export pump deficiency. *J Hepatol* 73:84, 2020.
- W AW: Organic anion uptake by hepatocytes. *Compr Physiol* 4:1715, 2014.

replicates like a retrovirus. Although these agents can be distinguished by their molecular and antigenic properties, all types of viral hepatitis produce clinically similar illnesses. These range from asymptomatic and inapparent to fulminant and fatal acute infections common to all types, on the one hand, and from subclinical persistent infections to rapidly progressive chronic liver disease with cirrhosis and even hepatocellular carcinoma, common to the bloodborne types (HBV, HCV, and HDV), on the other.

VIROLOGY AND ETIOLOGY

Hepatitis A HAV is a nonenveloped 27-nm, heat-, acid-, and ether-resistant, single-stranded, positive-sense RNA virus in the *Hepadnavirus* genus of the picornavirus family (Fig. 339-1). Quasi-enveloped virus particles encased in host plasma membrane-derived membranous vesicles circulate in the bloodstream. The virion contains four structural capsid polypeptides, designated VP1–VP4, as well as six nonstructural proteins, which are cleaved posttranslationally from the polyprotein product of a 7500-nucleotide genome. Inactivation of viral activity can be achieved by boiling for 1 min, by contact with formaldehyde and chlorine, or by ultraviolet irradiation. Despite nucleotide sequence variation of up to 20% among isolates of HAV and despite the recognition of six genotypes (three of which affect humans), all strains of this virus are immunologically indistinguishable and belong to one serotype. Human HAV can infect and cause hepatitis in chimpanzees, tamarins (marmosets), and several monkey species. Recently, a hepatotropic *Hepadnavirus* related to, and likely to have shared common evolutionary ancestry with, human HAV has been identified in several species of harbor seals, albeit without histologic evidence for liver injury or inflammation; HAV-like hepatoviruses have also been identified in small mammals, including bats and rodents. Hepatitis A has an incubation period of ~3–4 weeks. Its replication is limited to the liver, but the virus is present in the liver, bile, stools, and blood during the late incubation period and acute preicteric/presymptomatic phase of illness. Despite slightly longer persistence of virus in the liver, fecal shedding, viremia, and infectivity diminish rapidly once jaundice becomes apparent. Detection of HAV RNA by sensitive reverse transcription polymerase chain reaction assays has been reported to persist at low levels in stool, the liver, and serum for up to several months after acute illness; however, this does not correlate with persistent infectivity, probably because of the presence of neutralizing antibody. HAV can be cultivated reproducibly in vitro and in primate models.

Antibodies to HAV (anti-HAV) can be detected during acute illness when serum aminotransferase activity is elevated and fecal HAV shedding is still occurring. This early antibody response is predominantly of the IgM class and persists for several (~3) months, rarely for 6–12 months. During convalescence, however, anti-HAV of the IgG class becomes the predominant antibody (Fig. 339-2). Therefore, the diagnosis of hepatitis A is made during acute illness by demonstrating anti-HAV of the IgM class. After acute illness, anti-HAV of the IgG class remains detectable indefinitely, and patients with serum anti-HAV are immune to reinfection. Neutralizing antibody activity parallels the appearance of anti-HAV, and the IgG anti-HAV present in immune globulin accounts for the protection it affords against HAV infection.

Hepatitis B HBV is a DNA virus with a remarkably compact genomic structure; despite its small, circular, 3200-bp size, HBV DNA codes for four sets of viral products with a complex, multiparticle structure. HBV achieves its genomic economy by relying on an efficient strategy of encoding proteins from four overlapping genes: S, C, P, and X (Fig. 339-3), as detailed below. Once thought to be unique among viruses, HBV is now recognized as one of a family of animal viruses, hepadnaviruses (hepatotropic DNA viruses), and is classified as hepadnavirus type 1. Similar viruses infect certain species of woodchucks, ground and tree squirrels, and Pekin ducks, to mention the most carefully characterized; genetic evidence of ancient HBV-like virus forbears has been found in fossils of ancient birds, and an HBV-like virus has been identified in contemporary fish. Studies of ancient HBV genomes date an association between HBV and human beings back as long as 21,000 years ago; primate HBV-like viruses date back millions of years,

339

Acute Viral Hepatitis

Jules L. Dienstag



Acute viral hepatitis is a systemic infection affecting the liver predominantly. Almost all cases of acute viral hepatitis are caused by one of five viral agents: hepatitis A virus (HAV), hepatitis B virus (HBV), hepatitis C virus (HCV), the HBV-associated delta agent or hepatitis D virus (HDV), and hepatitis E virus (HEV). All these human hepatitis viruses are RNA viruses, except for hepatitis B, which is a DNA virus but

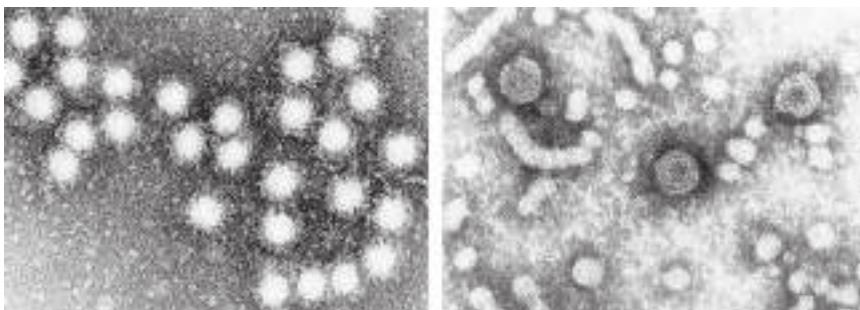


FIGURE 339-1 Electron micrographs of hepatitis A virus particles and serum from a patient with hepatitis B. *Left:* 27-nm hepatitis A virus particles purified from stool of a patient with acute hepatitis A and aggregated by antibody to hepatitis A virus. *Right:* Concentrated serum from a patient with hepatitis B, demonstrating the 42-nm virions, tubular forms, and spherical 22-nm particles of hepatitis B surface antigen. 132,000 \times . (Hepatitis D resembles 42-nm virions of hepatitis B but is smaller, 35–37 nm; hepatitis E resembles hepatitis A virus but is slightly larger, 32–34 nm; hepatitis C has been visualized as a 55-nm particle.)

suggesting that HBV predated the emergence of modern humans. Like HBV, all have the same distinctive three morphologic forms, have counterparts to the envelope and nucleocapsid virus antigens of HBV, replicate in the liver but exist in extrahepatic sites, contain their own endogenous DNA polymerase, have partially double-strand and partially single-strand genomes, are associated with acute and chronic hepatitis and hepatocellular carcinoma, and rely on a replicative strategy unique among DNA viruses but typical of retroviruses. Entry of HBV into hepatocytes is mediated by binding to the sodium taurocholate cotransporting polypeptide receptor. Instead of DNA replication directly from a DNA template, hepadnaviruses rely on reverse transcription (effected by the DNA polymerase) of minus-strand DNA from a “pregenomic” RNA intermediate. Then, plus-strand DNA is transcribed from the minus-strand DNA template by the DNA-dependent DNA polymerase and converted in the hepatocyte nucleus to a covalently closed circular DNA, which serves as a template for messenger RNA and pregenomic RNA. Viral proteins are translated by the messenger RNA, and the proteins and genome are packaged into virions and secreted from the hepatocyte. Although HBV is difficult to cultivate in vitro in the conventional sense from clinical material, several cell lines have been transfected with HBV DNA. Such transfected cells support in vitro replication of the intact virus and its component proteins.

VIRAL PROTEINS AND PARTICLES Of the three particulate forms of HBV (Table 339-1), the most numerous are the 22-nm particles, which appear as spherical or long filamentous forms; these are antigenically indistinguishable from the outer surface or envelope protein of HBV and are thought to represent excess viral envelope protein. Outnumbered in serum by a factor of 100 or 1000 to 1 compared with the spheres and tubules are large, 42-nm, double-shelled spherical particles, which represent the intact hepatitis B virion (Fig. 339-1). The envelope protein expressed on the outer surface of the virion and on the smaller spherical and tubular structures is referred to as *hepatitis*

B surface antigen (HBsAg). The concentration of HBsAg and virus particles in the blood may reach 500 μ g/mL and 10 trillion particles per milliliter, respectively. The envelope protein, HBsAg, is the product of the S gene of HBV.

Envelope HBsAg subdeterminants include a common group-reactive antigen, *a*, shared by all HBsAg isolates and one of several subtype-specific antigens—*d* or *y*, *w* or *r*—as well as other specificities. Hepatitis B isolates fall into one of at least 8 subtypes and 10 genotypes (A–J). Geographic distribution of genotypes and subtypes varies; genotypes A (corresponding to subtype *adw*) and D (*ayw*) predominate in the United States and Europe, whereas genotypes B (*adw*) and C (*adr*) predominate in Asia; how-

ever, these geographic distinctions have been blunted by recent-decade migration across continents. Clinical course and outcome are independent of subtype, but genotype B appears to be associated with less rapidly progressive liver disease and cirrhosis and a lower likelihood, or delayed appearance, of hepatocellular carcinoma than genotype C or D. Patients with genotype A are more likely to clear circulating viremia and achieve *hepatitis B e antigen* (HBeAg) and HBsAg seroconversion, both spontaneously and in response to antiviral therapy. In addition, “precore” mutations are favored by certain genotypes (see below).

Upstream of the S gene are the pre-S genes (Fig. 339-3), which code for pre-S gene products, including receptors on the HBV surface for polymerized human serum albumin and for hepatocyte membrane proteins. The pre-S region actually consists of both pre-S1 and pre-S2. Depending on where translation is initiated, three potential HBsAg gene products are synthesized. The protein product of the S gene is HBsAg (*major protein*), the product of the S region plus the adjacent pre-S2 region is the *middle protein*, and the product of the pre-S1 plus pre-S2 plus S regions is the *large protein*. Compared with the smaller spherical and tubular particles of HBV, complete 42-nm virions are enriched in the large protein. Both pre-S proteins and their respective antibodies can be detected during HBV infection, and the period of pre-S antigenemia appears to coincide with other markers of virus replication, as detailed below; however, pre-S proteins have little clinical relevance and are not included in routine serologic testing repertoires.

The intact 42-nm virion contains a 27-nm nucleocapsid core particle. Nucleocapsid proteins are coded for by the C gene. The antigen expressed on the surface of the nucleocapsid core is *hepatitis B core antigen* (HBcAg), and its corresponding antibody is anti-HBc. A third HBV antigen is HBeAg, a soluble, nonparticulate, nucleocapsid protein that is immunologically distinct from intact HBcAg but is a product of the same C gene. The C gene has two initiation codons: a precore and a core region (Fig. 339-3). If translation is initiated at the precore region, the protein product is HBeAg, which has a signal peptide that binds it to the smooth endoplasmic reticulum, the secretory apparatus of the cell, leading to its secretion into the circulation. If translation begins at the core region, HBcAg is the protein product; it has no signal peptide and is not secreted, but it assembles into nucleocapsid particles, which bind to and incorporate RNA, and which, ultimately, contain HBV DNA. Also packaged within the nucleocapsid core is a DNA polymerase, which directs replication and repair of HBV DNA. When packaging within viral proteins is complete, synthesis of the incomplete plus strand stops; this accounts for the single-strand gap and for differences in the size of the gap. HBcAg particles remain in the hepatocyte, where they are readily detectable by immunohistochemical staining and are exported after encapsidation by an envelope of HBsAg. Therefore, naked core particles do not circulate in the serum. The secreted nucleocapsid protein, HBeAg, provides a convenient, readily detectable, qualitative marker of HBV replication and relative infectivity.

HBsAg-positive serum containing HBeAg is more likely to be highly infectious and to be associated with the presence of hepatitis B

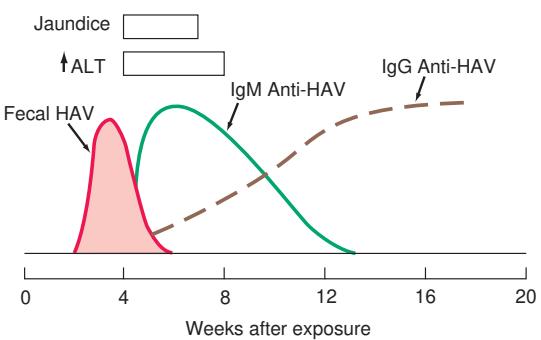


FIGURE 339-2 Scheme of typical clinical and laboratory features of hepatitis A virus (HAV). ALT, alanine aminotransferase.

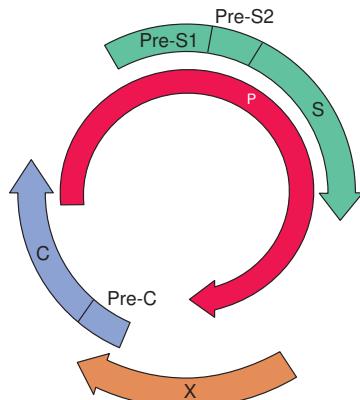


FIGURE 339-3 Compact genomic structure of hepatitis B virus (HBV). This structure, with overlapping genes, permits HBV to code for multiple proteins. The S gene codes for the “major” envelope protein, HBsAg. Pre-S1 and pre-S2, upstream of S, combine with S to code for two larger proteins, “middle” protein, the product of pre-S2 + S, and “large” protein, the product of pre-S1 + pre-S2 + S. The largest gene, P, codes for DNA polymerase. The C gene codes for two nucleocapsid proteins, HBeAg, a soluble, secreted protein (initiation from the pre-C region of the gene), and HBcAg, the intracellular core protein (initiation after pre-C). The X gene codes for HBxAg, which can transactivate the transcription of cellular and viral genes; its clinical relevance is not known, but it may contribute to carcinogenesis by binding to p53.

virions (and detectable HBV DNA, see below) than HBeAg-negative or anti-HBe-positive serum. For example, HBsAg-positive mothers who are HBeAg-positive almost invariably (>90%) transmit hepatitis B infection to their offspring, whereas HBsAg-positive mothers with anti-HBe rarely (10–15%) infect their offspring.

Early during the course of acute hepatitis B, HBeAg appears transiently; its disappearance may be a harbinger of clinical improvement and resolution of infection. Persistence of HBeAg in serum beyond the first 3 months of acute infection may be predictive of the development of chronic infection, and the presence of HBeAg during chronic hepatitis B tends to be associated with ongoing viral replication, infectivity, and inflammatory liver injury (except during the early decades after perinatally acquired HBV infection; see below).

The third and largest of the HBV genes, the P gene (Fig. 339-3), codes for HBV DNA polymerase; as noted above, this enzyme has both DNA-dependent DNA polymerase and RNA-dependent reverse transcriptase activities. The fourth gene, X, codes for a small, non-particulate protein, *hepatitis B x antigen* (HBxAg), that is capable of transactivating the transcription of both viral and cellular genes (Fig. 339-3). In the cytoplasm, HBxAg effects calcium release (possibly from mitochondria), which activates signal-transduction pathways that lead to stimulation of HBV reverse transcription and HBV DNA replication. Such transactivation may enhance the replication of HBV, leading to the clinical association observed between the expression of HBxAg and antibodies to it in patients with severe chronic

TABLE 339-1 Nomenclature and Features of Hepatitis Viruses

HEPATITIS TYPE	VIRUS PARTICLE, nm	MORPHOLOGY	GENOME ^a	CLASSIFICATION	ANTIGEN(S)	ANTIBODIES	REMARKS
HAV	27	Icosahedral nonenveloped	7.5-kb RNA, linear, ss, +	Hepadnavirus	HAV	Anti-HAV	Early fecal shedding Diagnosis: IgM anti-HAV Previous infection: IgG anti-HAV
HBV	42	Double-shelled virion (surface and core) spherical	3.2-kb DNA, circular, ss/ds	Hepadnavirus	HBsAg	Anti-HBs	Bloodborne virus; carrier state
	27	Nucleocapsid core			HBcAg	Anti-HBc	Acute diagnosis: HBsAg, IgM anti-HBc
	22	Spherical and filamentous; represents excess virus coat material			HBeAg	Anti-HBe	Chronic diagnosis: IgG anti-HBc, HBsAg Markers of replication: HBeAg, HBV DNA Liver, lymphocytes, other organs
HCV	55	Enveloped	9.4-kb RNA, linear, ss, +	Hepacivirus	HCV core antigen	Anti-HCV	Nucleocapsid contains DNA and DNA polymerase; present in hepatocyte nucleus; HBcAg does not circulate; HBeAg (soluble, nonparticulate) and HBV DNA circulate—correlate with infectivity and complete virions
					HBsAg	Anti-HBs	HBsAg detectable in >95% of patients with acute hepatitis B; found in serum, body fluids, hepatocyte cytoplasm; anti-HBs appears following infection—protective antibody
HDV	35–37	Enveloped hybrid particle with HBsAg coat and HDV core	1.7-kb RNA, circular, ss, –	Resembles viroids and plant satellite viruses (genus Deltavirus)	HBsAg	Anti-HBs	Bloodborne agent, formerly labeled non-A, non-B hepatitis
					HDAg	Anti-HDV	Acute diagnosis: anti-HCV, HCV RNA Chronic diagnosis: anti-HCV, HCV RNA; cytoplasmic location in hepatocytes
HEV	32–34	Nonenveloped icosahedral	7.6-kb RNA, linear, ss, +	Orthohepevirus	HEV antigen	Anti-HEV	Defective RNA virus, requires helper function of HBV (hepadnaviruses); HDV antigen (HDAg) present in hepatocyte nucleus
							Diagnosis: anti-HDV, HDV RNA; HBV/HDV co-infection—IgM anti-HBc and anti-HDV; HDV superinfection—IgG anti-HBc and anti-HDV
^a ss, single-strand; ss/ds, partially single-strand, partially double-strand; –, minus-strand; +, plus-strand.							

Note: See text for abbreviations.

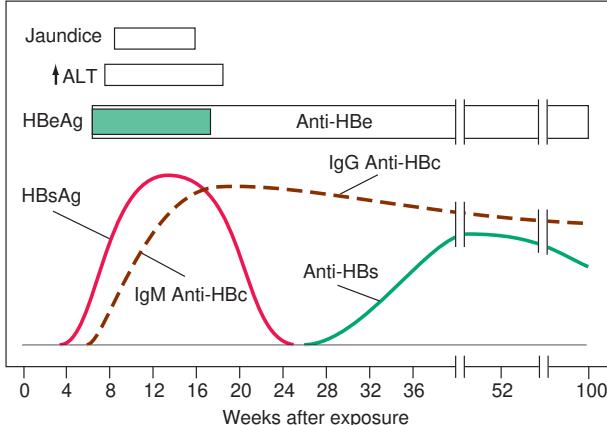


FIGURE 339-4 Scheme of typical clinical and laboratory features of acute hepatitis B. ALT, alanine aminotransferase.

hepatitis and hepatocellular carcinoma. The transactivating activity can enhance the transcription and replication of other viruses besides HBV, such as HIV. Cellular processes transactivated by X include the human interferon- γ gene and class I major histocompatibility genes; potentially, these effects could contribute to enhanced susceptibility of HBV-infected hepatocytes to cytolytic T cells. The expression of X can also induce programmed cell death (apoptosis). The clinical relevance of HBxAg is limited, however, and testing for it is not part of routine clinical practice.

SEROLOGIC AND VIROLOGIC MARKERS After a person is infected with HBV, the first virologic marker detectable in serum within 1–12 weeks, usually between 8 and 12 weeks, is HBsAg (Fig. 339-4). Circulating HBsAg precedes elevations of serum aminotransferase activity and clinical symptoms by 2–6 weeks and remains detectable during the entire icteric or symptomatic phase of acute hepatitis B and beyond. In typical cases, HBsAg becomes undetectable 1–2 months after the onset of jaundice and rarely persists beyond 6 months. After HBsAg disappears, antibody to HBsAg (anti-HBs) becomes detectable in serum and remains detectable indefinitely thereafter. Because HBcAg is intracellular and, when in the serum, sequestered within an HBsAg coat, naked core particles do not circulate in serum, and therefore, HBcAg is not detectable routinely in the serum of patients with HBV infection. By contrast, anti-HBc is readily demonstrable in serum, beginning within the first 1–2 weeks after the appearance of HBsAg and preceding detectable levels of anti-HBs by weeks to months. Because variability exists in the time of appearance of anti-HBs after HBV infection, occasionally a gap of several weeks or longer may separate the disappearance of HBsAg and the appearance of anti-HBs. During this “gap” or “window” period, anti-HBc may represent the only serologic evidence of current or recent HBV infection, and blood containing anti-HBc in the absence of HBsAg and anti-HBs has been implicated in transfusion-associated hepatitis B. In part because the sensitivity of immunoassays for HBsAg and anti-HBs has increased, however, this window period is rarely encountered. In some persons, years after HBV infection, anti-HBc may persist in the circulation longer than anti-HBs. Therefore, isolated anti-HBc does not necessarily indicate active virus replication; most instances of isolated anti-HBc represent hepatitis B infection in the remote past. Rarely, however, isolated anti-HBc represents low-level hepatitis B viremia, with HBsAg below the detection threshold, and occasionally, isolated anti-HBc represents a cross-reacting or false-positive immunologic specificity. Recent and remote HBV infections can be distinguished by determination of the immunoglobulin class of anti-HBc. Anti-HBc of the IgM class (IgM anti-HBc) predominates during the first 6 months after acute infection, whereas IgG anti-HBc is the predominant class of anti-HBc beyond 6 months. Therefore, patients with current or recent acute hepatitis B, including those in the anti-HBc window, have IgM anti-HBc in their serum. In patients who have recovered from hepatitis B in the remote past as well

as those with chronic HBV infection, anti-HBc is predominantly of the IgG class. Infrequently, in <1–5% of patients with acute HBV infection, levels of HBsAg are too low to be detected; in such cases, the presence of IgM anti-HBc establishes the diagnosis of acute hepatitis B. When isolated anti-HBc occurs in the rare patient with chronic hepatitis B whose HBsAg level is below the sensitivity threshold of contemporary immunoassays (a low-level carrier), anti-HBc is of the IgG class. Generally, in persons who have recovered from hepatitis B, anti-HBs and anti-HBc persist indefinitely.

The temporal association between the appearance of anti-HBs and resolution of HBV infection as well as the observation that persons with anti-HBs in serum are protected against reinfection with HBV suggests that *anti-HBs is the protective antibody*. Therefore, strategies for prevention of HBV infection are based on providing susceptible persons with circulating anti-HBs (see below). Occasionally, in ~10% of patients with chronic hepatitis B, low-level, low-affinity anti-HBs can be detected. This antibody is directed against a subtype determinant different from that represented by the patient's HBsAg; its presence is thought to reflect the stimulation of a related clone of antibody-forming cells, but it has no clinical relevance and does not signal imminent clearance of hepatitis B. These patients with HBsAg and such nonneutralizing anti-HBs should be categorized as having chronic HBV infection.

The other readily detectable serologic marker of HBV infection, HBeAg, appears concurrently with or shortly after HBsAg. Its appearance coincides temporally with high levels of virus replication and reflects the presence of circulating intact virions and detectable HBV DNA (with the notable exception of patients with precore mutations who cannot synthesize HBeAg—see “Molecular Variants”). Pre-S1 and pre-S2 proteins are also expressed during periods of peak replication, but assays for these gene products are not routinely available. In self-limited HBV infections, HBeAg becomes undetectable shortly after peak elevations in aminotransferase activity, before the disappearance of HBsAg, and anti-HBc then becomes detectable, coinciding with a period of relatively lower infectivity (Fig. 339-4). Because markers of HBV replication appear transiently during acute infection, testing for such markers is of little clinical utility in typical cases of acute HBV infection. In contrast, markers of HBV replication provide valuable information in patients with protracted infections.

Departing from the pattern typical of acute HBV infections, in chronic HBV infection, HBsAg remains detectable beyond 6 months, anti-HBc is primarily of the IgG class, and anti-HBs is either undetectable or detectable at low levels (see “Laboratory Features”) (Fig. 339-5).

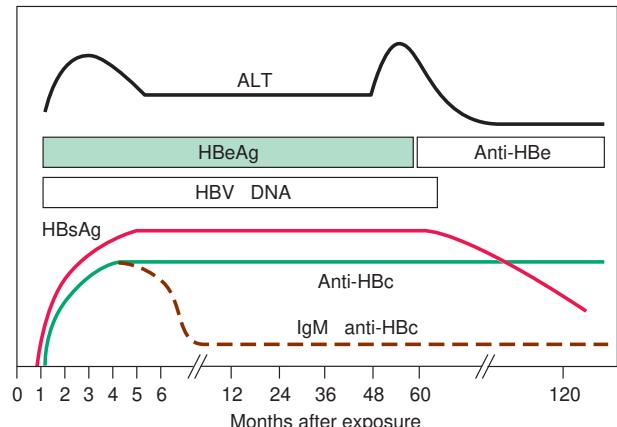


FIGURE 339-5 Scheme of typical laboratory features of wild-type chronic hepatitis B. HBeAg and hepatitis B virus (HBV) DNA can be detected in serum during the relatively *replicative phase* of chronic infection, which is associated with infectivity and liver injury. Seroconversion from the replicative phase to the relatively *nonreplicative phase* occurs at a rate of ~10% per year and is heralded by an acute hepatitis-like elevation of alanine aminotransferase (ALT) activity; during the nonreplicative phase, infectivity and liver injury are limited. In HBeAg-negative chronic hepatitis B associated with mutations in the precore region of the HBV genome, replicative chronic hepatitis B occurs in the absence of HBeAg.

During early chronic HBV infection, HBV DNA can be detected both in serum and in hepatocyte nuclei, where it is present in free or episomal form. This relatively highly *replicative* stage of HBV infection is the time of maximal infectivity and liver injury; HBeAg is a qualitative marker and HBV DNA a quantitative marker of this replicative phase, during which all three forms of HBV circulate, including intact virions. Over time, the relatively replicative phase of chronic HBV infection gives way to a relatively *nonreplicative phase*. This occurs at a rate of ~10% per year and is accompanied by seroconversion from HBeAg to anti-HBe. In many cases, this seroconversion coincides with a transient, usually mild, acute hepatitis-like elevation in aminotransferase activity, believed to reflect cell-mediated immune clearance of virus-infected hepatocytes. In this relatively nonreplicative phase of chronic infection, when HBV DNA is demonstrable in hepatocyte nuclei, it tends to be integrated into the host genome. In this phase, only spherical and tubular forms of HBV, *not intact virions*, circulate, and liver injury tends to subside. Most such patients would be characterized as *inactive HBV carriers*. In reality, the designations *replicative* and *nonreplicative* are only relative; even in the so-called nonreplicative phase, HBV replication can be detected at levels of approximately $\leq 10^3$ virions/mL with highly sensitive amplification probes such as the polymerase chain reaction (PCR); below this replication threshold, liver injury and infectivity of HBV are limited to negligible. Still, the distinctions are pathophysiologically and clinically meaningful. Occasionally, nonreplicative HBV infection converts back to replicative infection. Such spontaneous reactivations are accompanied by reexpression of HBeAg and HBV DNA, and sometimes of IgM anti-HBc, as well as by exacerbations of liver injury. Because high-titer IgM anti-HBc can reappear during acute exacerbations of chronic hepatitis B, relying on IgM anti-HBc versus IgG anti-HBc to distinguish between acute and chronic hepatitis B infection, respectively, may not always be reliable; in such cases, patient history and additional follow-up monitoring over time are invaluable in helping to distinguish de novo acute hepatitis B infection from acute exacerbation of chronic hepatitis B infection.

MOLECULAR VARIANTS Variation occurs throughout the HBV genome, and clinical isolates of HBV that do not express typical viral proteins have been attributed to mutations in individual or even multiple gene locations. For example, variants have been described that lack nucleocapsid proteins (commonly), envelope proteins (very rarely), or both. Two categories of naturally occurring HBV variants have attracted the most attention. One of these was identified initially in Mediterranean countries among patients with severe chronic HBV infection and detectable HBV DNA, but with anti-HBe instead of HBeAg. These patients were found to be infected with an HBV mutant that contained an alteration in the precore region, rendering the virus incapable of encoding HBeAg. Although several potential mutation sites exist in the pre-C region, the region of the C gene necessary for the expression of HBeAg (see "Virology and Etiology"), the most commonly encountered in such patients is a single base substitution, from G to A in the second to last codon of the pre-C gene at nucleotide 1896. This substitution results in the replacement of the TGG tryptophan codon by a stop codon (TAG), which prevents the translation of HBeAg. Another mutation, in the core-promoter region, prevents transcription of the coding region for HBeAg and yields an HBeAg-negative phenotype. Patients with such mutations in the precore region and who are unable to secrete HBeAg may have severe liver disease that progresses more rapidly to cirrhosis, or alternatively, they are identified clinically later in the course of the natural history of chronic hepatitis B, when the disease is more advanced. Both "wild-type" HBV and precore-mutant HBV can coexist in the same patient, or mutant HBV may arise late during wild-type HBV infection. In addition, clusters of fulminant hepatitis B in Israel and Japan were attributed to common-source infection with a precore mutant. Fulminant hepatitis B in North America and western Europe, however, occurs in patients infected with wild-type HBV, in the absence of precore mutants, and both precore mutants and other mutations throughout the HBV genome occur commonly, even in patients with typical, self-limited, milder forms of HBV infection. HBeAg-negative

chronic hepatitis with mutations in the precore region is now the most frequently encountered form of hepatitis B in Mediterranean countries and in Europe. In the United States, where HBV genotype A (less prone to G1896A mutation) is prevalent, precore-mutant HBV is much less common; however, as a result of immigration from Asia and Europe, the proportion of HBeAg-negative hepatitis B-infected individuals has increased in the United States, and they now represent ~30–40% of patients with chronic hepatitis B. Characteristic of such HBeAg-negative chronic hepatitis B are lower levels of HBV DNA (usually $\leq 10^5$ IU/mL) and one of several patterns of aminotransferase activity—persistent elevations, periodic fluctuations above the normal range, and periodic fluctuations between the normal and elevated range.

The second important category of HBV mutants consists of *escape mutants*, in which a single amino acid substitution, from glycine to arginine, occurs at position 145 of the immunodominant *a* determinant common to all HBsAg subtypes. This HBsAg alteration leads to a critical conformational change that results in a loss of neutralizing activity by anti-HBs. This specific HBV/a mutant has been observed in two situations, active and passive immunization, in which humoral immunologic pressure may favor evolutionary change ("escape") in the virus—in a small number of hepatitis B vaccine recipients who acquired HBV infection despite the prior appearance of neutralizing anti-HBs and in HBV-infected liver transplant recipients treated with a high-potency human monoclonal anti-HBs preparation. Although such mutants have not been recognized frequently, their existence raises a concern that may complicate vaccination strategies and serologic diagnosis.

Different types of mutations emerge during antiviral therapy of chronic hepatitis B with nucleoside analogues; such YMDD and similar mutations in the polymerase motif of HBV are described in Chap. 341.

EXTRAHEPATIC SITES Hepatitis B antigens and HBV DNA have been identified in extrahepatic sites, including the lymph nodes, bone marrow, circulating lymphocytes, spleen, and pancreas. Although the virus does not appear to be associated with tissue injury in any of these extrahepatic sites, its presence in these "remote" reservoirs has been invoked (but is not necessary) to explain the recurrence of HBV infection after orthotopic liver transplantation. The clinical relevance of such extrahepatic HBV is limited.

Hepatitis D The delta hepatitis agent, or HDV, the only member of the genus *Deltavirus*, is a defective RNA virus that co-infects with and requires the helper function of HBV (or other hepadnaviruses) for its replication and expression. Slightly smaller than HBV, HDV is a formalin-sensitive, 35- to 37-nm virus with a hybrid structure. Its nucleocapsid expresses HDV antigen (HDAG), which bears no antigenic homology with any of the HBV antigens, and contains the virus genome. The HDV core is "encapsidated" by an outer envelope of HBsAg, indistinguishable from that of HBV except in its relative compositions of major, middle, and large HBsAg component proteins. The genome is a small, 1700-nucleotide, circular, single-strand RNA of negative polarity that is nonhomologous with HBV DNA (except for a small area of the polymerase gene) but that has features and the rolling circle model of replication common to genomes of plant satellite viruses or viroids. HDV RNA contains many areas of internal complementarity; therefore, it can fold on itself by internal base pairing to form an unusual, very stable, rod-like structure that contains a very stable, self-cleaving and self-ligating ribozyme. HDV RNA requires host RNA polymerase II for its replication in the hepatocyte nucleus via RNA-directed RNA synthesis by transcription of genomic RNA to a complementary antigenomic (plus strand) RNA; the antigenomic RNA, in turn, serves as a template for subsequent genomic RNA synthesis effected by host RNA polymerase I. HDV RNA has only one open reading frame, and HDAG, a product of the antigenomic strand, is the only known HDV protein; HDAG exists in two forms: a small, 195-amino-acid species, which plays a role in facilitating HDV RNA replication, and a large, 214-amino-acid species, which appears to suppress replication but is required for assembly of the antigen into virions. HDV antigens have been shown to bind directly to RNA

polymerase II, resulting in stimulation of transcription. Viral assembly requires farnesylation of the large HDAg for ribonucleoprotein anchoring to HBsAg. Both HBV and HDV enter hepatocytes via the sodium taurocholate cotransporting polypeptide receptor. Although complete hepatitis D virions and liver injury require the cooperative helper function of HBV, intracellular replication of HDV RNA can occur without HBV. Genomic heterogeneity among HDV isolates has been described. Although pathophysiologic and clinical consequences of this genetic diversity have not been defined definitively, preliminarily, genotype 2 has been linked to milder disease and genotype 3 to severe acute disease. The clinical spectrum of hepatitis D is common to all eight genotypes identified, the predominant of which is genotype 1.

HDV can either infect a person simultaneously with HBV (*co-infection*) or superinfect a person already infected with HBV (*superinfection*); when HDV infection is transmitted from a donor with one HBsAg subtype to an HBsAg-positive recipient with a different subtype, HDV assumes the HBsAg subtype of the recipient, rather than the donor. Because HDV relies absolutely on HBV, the duration of HDV infection is determined by the duration of (and cannot outlast) HBV infection. HDV replication tends to suppress HBV replication; therefore, patients with hepatitis D tend to have lower levels of HBV replication. HDV antigen is expressed primarily in hepatocyte nuclei and is occasionally detectable in serum. During acute HDV infection, anti-HDV of the IgM class predominates, and 30–40 days may elapse after symptoms appear before anti-HDV can be detected. In self-limited infection, anti-HDV is low-titer and transient, rarely remaining detectable beyond the clearance of HBsAg and HDV antigen. In chronic HDV infection, anti-HDV circulates in high titer, and both IgM and IgG anti-HDV can be detected. HDV antigen in the liver and HDV RNA in serum and liver can be detected during HDV replication.

The recent report that, *in vitro*, HDV can assemble infectious virus particles with envelope glycoproteins from other viruses, both hepatotropic and nonhepatotropic, raises the possibility that HDV can replicate without hepadnaviruses; however, to date, co-infections in nature with other viruses have not been observed.

Hepatitis C Hepatitis C virus, which, before its identification, was labeled “non-A, non-B hepatitis,” is a linear, single-strand, positive-sense, 9600-nucleotide RNA virus, the genome of which is similar in organization to that of flaviviruses and pestiviruses; HCV is the only member of the genus *Hepacivirus* in the family Flaviviridae. The HCV genome contains a single, large open reading frame (ORF) (gene) that codes for a virus polyprotein of ~3000 amino acids, which is cleaved after translation to yield 10 viral proteins. The 5' end of the genome consists of an untranslated region (containing an internal ribosomal entry site [IRES]) adjacent to the genes for three structural proteins, the nucleocapsid core protein, C, and two envelope glycoproteins, E1 and E2. The 5' untranslated region and core gene are highly conserved among genotypes, but the envelope proteins are coded for by the

hypervariable region, which varies from isolate to isolate and may allow the virus to evade host immunologic containment directed at accessible virus-envelope proteins. The 3' end of the genome also includes an untranslated region and contains the genes for seven nonstructural (NS) proteins: p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B. p7 is a membrane ion channel protein necessary for efficient assembly and release of HCV. The NS2 cysteine protease cleaves NS3 from NS2, and the NS3-4A serine protease cleaves all the downstream proteins from the polyprotein. Important NS proteins involved in virus replication include the NS3 helicase; NS3-4A serine protease; the multifunctional membrane-associated phosphoprotein NS5A, an essential component of the viral replication membranous web (along with NS4B); and the NS5B RNA-dependent RNA polymerase (Fig. 339-6). Because HCV does not replicate via a DNA intermediate, it does not integrate into the host genome. Because HCV tends to circulate in relatively low titer, 10³–10⁷ virions/mL, visualization of the 50- to 80-nm virus particles remains difficult. Still, the replication rate of HCV is very high, 10¹² virions per day; its half-life is 2.7 h. The chimpanzee is a helpful but cumbersome animal model. Although a robust, reproducible, small animal model is lacking, HCV replication has been documented in an immunodeficient mouse model containing explants of human liver and in transgenic mouse and rat models; in addition, an HCV-related rat *Hepacivirus* has been reported to be a useful surrogate model. Although *in vitro* replication is difficult, replicons in hepatocellular carcinoma-derived cell lines support replication of genetically manipulated, truncated, or full-length HCV RNA (but not intact virions); infectious pseudotyped retroviral HCV particles have been shown to yield functioning envelope proteins. In 2005, complete replication of HCV and intact 55-nm virions were described in cell culture systems. HCV entry into the hepatocyte occurs via the non-liver-specific CD81 receptor and the liver-specific tight junction protein claudin-1. A growing list of additional host receptors to which HCV binds on cell entry includes occludin, low-density lipoprotein receptors, glycosaminoglycans, scavenger receptor B1, and epidermal growth factor receptor, among others. Relying on the same assembly and secretion pathway as low-density and very-low-density lipoproteins, HCV is a lipoprotein particle and masquerades as a lipoprotein, which may limit its visibility to the adaptive immune system and explain its ability to evade immune containment and clearance. After viral entry and uncoating, translation is initiated by the IRES on the endoplasmic reticulum membrane, and the HCV polyprotein is cleaved during translation and posttranslationally by host cellular proteases as well as HCV NS2-3 and NS3-4A proteases. Host cofactors involved in HCV replication include cyclophilin A, which binds to NS5A and yields conformational changes required for viral replication, and liver-specific host microRNA miR-122.

At least six distinct major genotypes (and a minor genotype 7), as well as >50 subtypes within genotypes, of HCV have been identified by nucleotide sequencing. Genotypes differ from one another in sequence homology by ≥30%, and subtypes differ by ~20%. Because divergence of HCV isolates within a genotype or subtype and within the same host may vary insufficiently to define a distinct genotype, these intragenotypic differences are referred to as *quasispecies* and differ in sequence homology by only a few percent. The genotypic and quasispecies diversity of HCV, resulting from its high mutation rate, interferes with effective humoral immunity. Neutralizing antibodies to HCV have been demonstrated, but they tend to be short-lived, and HCV infection does not induce lasting immunity against reinfection with different virus isolates or even the same virus isolate. Thus, neither *heterologous* nor *homologous* immunity appears to develop commonly after acute HCV infection. Some HCV genotypes are distributed worldwide,

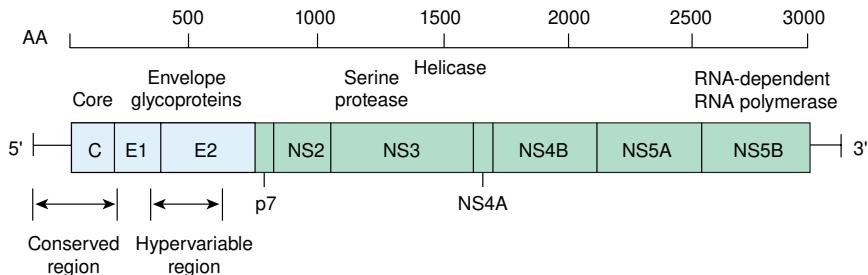


FIGURE 339-6 Organization of the hepatitis C virus genome and its associated, 3000-amino-acid (AA) proteins. The three structural genes at the 5' end are the core region, C, which codes for the nucleocapsid, and the envelope regions, E1 and E2, which code for envelope glycoproteins. The 5' untranslated region and the C region are highly conserved among isolates, whereas the envelope domain E2 contains the hypervariable region. At the 3' end are seven nonstructural (NS) regions—p7, a membrane protein adjacent to the structural proteins that appears to function as an ion channel; NS2, which codes for a cysteine protease; NS3, which codes for a serine protease and an RNA helicase; NS4 and NS4B; NS5A, a multifunctional membrane-associated phosphoprotein, an essential component of the viral replication membranous web; and NS5B, which codes for an RNA-dependent RNA polymerase. After translation of the entire polyprotein, individual proteins are cleaved by both host and viral proteases.

whereas others are more geographically confined (see “Epidemiology and Global Features”). In addition, differences exist among genotypes in responsiveness to antiviral therapy but not in pathogenicity or clinical progression (except for genotype 3, in which hepatic steatosis and clinical progression are more likely).

Currently available, third-generation immunoassays, which incorporate proteins from the core, NS3, and NS5 regions, detect anti-HCV antibodies during acute infection. The most sensitive indicator of HCV infection is the presence of HCV RNA, which requires molecular amplification by PCR or transcription-mediated amplification (TMA) (Fig. 339-7). To allow standardization of the quantification of HCV RNA among laboratories and commercial assays, HCV RNA is reported as international units (IUs) per milliliter; quantitative assays with a broad dynamic range are available that allow detection of HCV RNA with a sensitivity as low as 5 IU/mL. HCV RNA can be detected within a few days of exposure to HCV—well before the appearance of anti-HCV—and tends to persist for the duration of HCV infection. Application of sensitive molecular probes for HCV RNA has revealed the presence of replicative HCV in peripheral blood lymphocytes of infected persons; however, as is the case for HBV in lymphocytes, the clinical relevance of HCV lymphocyte infection is not known.

Hepatitis E Previously labeled *epidemic* or *enterically transmitted non-A, non-B hepatitis*, HEV is an enterically transmitted virus that causes clinically apparent hepatitis primarily in India, Asia, Africa, and Central America; in those geographic areas, HEV is the most common cause of acute hepatitis; one-third of the global population appears to have been infected. This agent, with epidemiologic features resembling those of hepatitis A, is a 27- to 34-nm, nonenveloped, heat-stable, HAV-like virus with a 7200-nucleotide, single-strand, positive-sense RNA genome. Like HAV, HEV also exists in a quasi-enveloped form enclosed within host-cell-derived membranes. HEV has three overlapping ORFs (genes), the largest of which, *ORF1*, encodes nonstructural proteins involved in virus replication (the viral replicase, which includes a protease, polymerase, and helicase). A middle-sized gene, *ORF2*, encodes the nucleocapsid protein, the major structural protein, and the smallest, *ORF3*, encodes a small structural phosphoprotein involved in virus particle secretion. All HEV isolates appear to belong to a single serotype, despite genomic heterogeneity of up to 25% and the existence of four species (A–D) and eight genotypes, only four of which, all within species A, have been detected in humans; genotypes 1 and 2 (common in developing countries) appear to be more virulent anthropotropic variants, whereas genotypes 3 (the most common in the United States and Europe) and 4 (seen in China), endemic in animal species (enzootic variants), are more attenuated, account for subclinical infections, represent a zoonotic reservoir for human infections, and can cause chronic infection in immunocompromised hosts. Contributing to the perpetuation of this virus are the animal reservoirs described above, most notably in swine but also in camels, deer, rats, and rabbits, among others. No genomic or antigenic homology, however, exists between HEV and HAV or other picornaviruses; and HEV, although resembling caliciviruses, is sufficiently distinct from any known agent

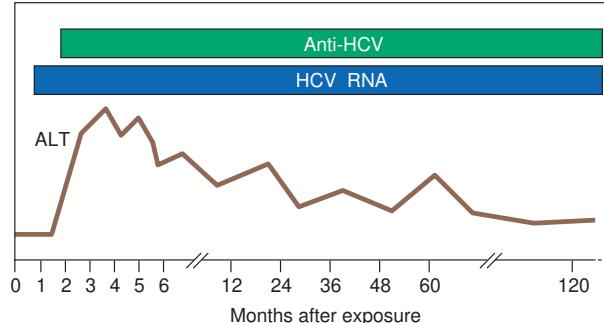


FIGURE 339-7 Scheme of typical laboratory features during acute hepatitis C progressing to chronicity. Hepatitis C virus (HCV) RNA is the first detectable event, preceding alanine aminotransferase (ALT) elevation and the appearance of anti-HCV.

to merit its own classification as a unique genus, *Orthohepevirus*, within the family Hepeviridae (which includes similar viruses infecting mammals, birds, and fish). The virus has been detected in stool, bile, and liver and is excreted in the stool during the late incubation period. Both IgM anti-HEV during early acute infection and IgG anti-HEV predominating after the first 3 months can be detected. The presence of HEV RNA in serum and stool accompanies acute infection; viremia resolves as clinical-biochemical recovery ensues, while HEV RNA in stool may outlast viremia by several weeks. Currently, serologic/virologic testing for HEV infection—not approved or licensed by the U.S. Food and Drug Administration (FDA)—can be done in specialized laboratories (e.g., the Centers for Disease Control and Prevention [CDC]) and some commercial laboratories.

■ PATHOGENESIS

Under ordinary circumstances, none of the hepatitis viruses is known to be directly cytopathic to hepatocytes. Evidence suggests that the clinical manifestations and outcomes after acute liver injury associated with viral hepatitis are determined by the immunologic responses of the host. Among the viral hepatides, the immunopathogenesis of hepatitis B and C has been studied most extensively.

Hepatitis B For HBV, the existence of inactive hepatitis B carriers with normal liver histology and function suggests that the virus is not directly cytopathic. The fact that patients with defects in cellular immune competence are more likely to remain chronically infected rather than to clear HBV supports the role of cellular immune responses in the pathogenesis of hepatitis B-related liver injury. The model that has the most experimental support involves cytolytic T cells sensitized specifically to recognize host and hepatitis B viral antigens on the liver cell surface. Nucleocapsid proteins (HBcAg and possibly HBeAg), present on the cell membrane in minute quantities, are the viral target antigens that, with host antigens, invite cytolytic T cells to destroy HBV-infected hepatocytes. Differences in the robustness and broad polyclonality of CD8+ cytolytic T-cell responsiveness; in the level of HBV-specific helper CD4+ T cells; in attenuation, depletion, and exhaustion of virus-specific T cells; in viral T-cell epitope escape mutations that allow the virus to evade T-cell containment; and in the elaboration of antiviral cytokines by T cells have been invoked to explain differences in outcomes between those who recover after acute hepatitis and those who progress to chronic hepatitis or between those with mild and those with severe (fulminant) acute HBV infection.

Although a robust cytolytic T-cell response occurs and eliminates virus-infected liver cells during acute hepatitis B, >90% of HBV DNA has been found in experimentally infected chimpanzees to disappear from the liver and blood before maximal T-cell infiltration of the liver and before most of the biochemical and histologic evidence of liver injury. This observation suggests that components of the innate immune system and inflammatory cytokines, independent of cytopathic antiviral mechanisms, participate in the early immune response to HBV infection; this effect has been shown to represent elimination of HBV replicative intermediates from the cytoplasm and covalently closed circular viral DNA from the nucleus of infected hepatocytes. In turn, the innate immune response to HBV infection is mediated largely by natural killer (NK) cell cytotoxicity, activated by immunosuppressive cytokines (e.g., interleukin [IL] 10 and transforming growth factor [TGF] β), reduced signals from inhibitory receptor expression (e.g., major histocompatibility complex), or increased signals from activating receptor expression on infected hepatocytes. In addition, NK cells reduce helper CD4+ cells, which results in reduced CD8+ cells and exhaustion of the virus-specific T-cell response to HBV infection. Adding to the evidence supporting the role of these immunologic perturbations in the pathogenesis of HBV-associated liver injury are the observations that many of these departures from normal immune function are restored after successful antiviral therapy. Ultimately, HBV-HLA-specific cytolytic T-cell responses of the adaptive immune system are felt to be responsible for recovery from HBV infection.

Debate continues over the relative importance of viral and host factors in the pathogenesis of HBV-associated liver injury and its outcome.

As noted above, precore genetic mutants of HBV have been associated with the more severe outcomes of HBV infection (severe chronic and fulminant hepatitis), suggesting that, under certain circumstances, relative pathogenicity is a property of the virus, not the host. The facts that concomitant HDV and HBV infections are associated with more severe liver injury than HBV infection alone and that cells transfected in vitro with the gene for HDV antigen express HDV antigen and then become necrotic in the absence of any immunologic influences are also consistent with a viral effect on pathogenicity. Similarly, in patients who undergo liver transplantation for end-stage chronic hepatitis B, occasionally, rapidly progressive liver injury appears in the new liver. This clinical pattern is associated with an unusual histologic pattern in the new liver, *fibrosing cholestatic hepatitis*, which, ultrastructurally, appears to represent a choking of the cell with overwhelming quantities of HBsAg. This observation suggests that, under the influence of the potent immunosuppressive agents required to prevent allograft rejection, HBV may have a direct cytopathic effect on liver cells, independent of the immune system.

Although the precise mechanism of liver injury in HBV infection remains elusive, studies of nucleocapsid proteins have shed light on the profound immunologic tolerance to HBV of babies born to mothers with highly replicative (HBeAg-positive), chronic HBV infection. In HBeAg-expressing transgenic mice, *in utero* exposure to HBeAg, which is sufficiently small to traverse the placenta, induces T-cell tolerance to both nucleocapsid proteins. This, in turn, may explain why, when infection occurs so early in life, immunologic clearance does not occur, and protracted, lifelong infection ensues. An alternative explanation proposed to explain why robust liver injury does not accompany neonatal HBV infection but predisposes to chronic infection is defective priming of HBV-specific T cells during *in utero* exposure to HBV.

"IMMUNOTOLERANT" VERSUS "IMMUNOREACTIVE" CHRONIC HEPATITIS B An important distinction should be drawn between HBV infection acquired at birth, common in endemic areas, such as East Asia, and infection acquired in adulthood, common in the West. Infection in the neonatal period is associated with the acquisition of what appears to be a high level of immunologic tolerance to HBV and absence of an acute hepatitis illness but the almost invariable establishment of chronic, often lifelong infection. Neonatally acquired HBV infection can culminate decades later in cirrhosis and hepatocellular carcinoma (see "Complications and Sequelae"). In contrast, when HBV infection is acquired during adolescence or early adulthood, the host immune response to HBV-infected hepatocytes tends to be robust, an acute hepatitis-like illness is the rule, and failure to recover is the exception. After adulthood-acquired infection, chronicity is uncommon, and the risk of hepatocellular carcinoma is very low. Based on these observations, some authorities categorize HBV infection into an "immunotolerant" phase, an "immunoreactive" phase, and an "inactive" phase. This somewhat simplistic formulation does not apply at all to the typical adult in the West with self-limited acute hepatitis B, in whom no period of immunologic tolerance occurs. Even among those with neonatally acquired HBV infection, in whom immunologic tolerance appears to be established definitively, immunologic responses to HBV infection have been demonstrated (albeit typically at reduced levels), and intermittent bursts of hepatic necroinflammatory activity punctuate the early decades of life during which liver injury appears to be quiescent (labeled by some as the "immunotolerant" phase; however, it more accurately is a period of dissociation between high-level HBV replication and a paucity of inflammatory liver injury). In addition, even when clinically apparent liver injury and progressive fibrosis emerge during later decades (the so-called immunoreactive, or immunointolerant, phase), the level of immunologic tolerance to HBV remains substantial. More accurately, in patients with neonatally acquired HBV infection, a dynamic equilibrium exists between tolerance and intolerance, the outcome of which determines the clinical expression of chronic infection. Persons infected as neonates tend to have a relatively higher level of immunologic tolerance (high replication, low necroinflammatory activity) during the early decades of life and a relatively lower level (but only rarely a loss) of tolerance (and

necroinflammatory activity reflecting the level of virus replication) in the later decades of life.

Hepatitis C Cell-mediated immune responses and elaboration by T cells of antiviral cytokines contribute to the multicellular innate and adaptive immune responses involved in the containment of infection and pathogenesis of liver injury associated with hepatitis C. The fact that HCV is so efficient in evading these immune mechanisms is a testament to its highly evolved ability to disrupt host immune responses at multiple levels. After exposure to HCV, the host cell identifies viral product motifs (pattern recognition receptors) that distinguish the virus from "self," resulting in the elaboration of interferons and other cytokines that result in activation of innate and adaptive immune responses. Intrahepatic human leukocyte antigen (HLA) class I-restricted cytolytic T cells directed at nucleocapsid, envelope, and nonstructural viral protein antigens have been demonstrated in patients with chronic hepatitis C; however, such virus-specific cytolytic T-cell responses do not correlate adequately with the degree of liver injury or with recovery. Yet a consensus has emerged supporting a role in the pathogenesis of HCV-associated liver injury of virus-activated CD4+ helper T cells that stimulate, via the cytokines they elaborate, HCV-specific CD8+ cytotoxic T cells. These responses appear to be more robust (higher in number, more diverse in viral antigen specificity, more functionally effective, and longer lasting) in those who recover from HCV infection than in those who have chronic infection. Contributing to chronic infection are a CD4+ proliferative defect that results in rapid contraction of CD4+ responses, mutations in CD8+ T cell-targeted viral epitopes that allow HCV to escape immune-mediated clearance, and upregulation of inhibitory receptors on functionally impaired, exhausted T cells. Although attention has focused on adaptive immunity, HCV proteins have been shown to interfere with innate immunity by resulting in blocking of type I interferon responses and inhibition of interferon signaling and effector molecules in the interferon signaling cascade.

Several HLA alleles have been linked with self-limited hepatitis C, the most convincing of which is the CC haplotype of the *IL28B* gene, which codes for interferon λ3, a component of innate immune antiviral defense. The *IL28B* association is even stronger when combined with HLA class II *DQ_{B1} 03:01*. The link between non-CC *IL28B* polymorphisms and failure to clear HCV infection has been explained by a chromosome 19q13.13 frameshift variant upstream of *IL28B*, the ΔG polymorphism of which creates an ORF in a novel interferon gene (*IFN-λ4*) associated with impaired HCV clearance. Also shown to contribute to limiting HCV infection are NK cells of the innate immune system that function when HLA class I molecules required for successful adaptive immunity are underexpressed. Both peripheral cytotoxicity and intrahepatic NK cell cytotoxicity are dysfunctional in persistent HCV infection. Adding to the complexity of the immune response, HCV core, NS4B, and NS5B have been shown to suppress the immunoregulatory nuclear factor (NF)-κB pathway, resulting in reduced antiapoptotic proteins and a resultant increased vulnerability to tumor necrosis factor (TNF) α-mediated cell death. Patients with hepatitis C and unfavorable (non-CC, associated with reduced HCV clearance) *IL28B* alleles have been shown to have depressed NK cell/innate immune function. Of note, the emergence of substantial viral quasispecies diversity and HCV sequence variation allow the virus to evade attempts by the host to contain HCV infection by both humoral and cellular immunity.

Finally, cross-reactivity between viral antigens (HCV NS3 and NS5A) and host autoantigens (cytochrome P450 2D6) has been invoked to explain the association between hepatitis C and a subset of patients with autoimmune hepatitis and antibodies to liver-kidney microsomal (LKM) antigen (anti-LKM) (Chap. 341).

Hepatitis A and E Viral shedding in these acute hepatitides predates clinical evidence of liver injury, consistent with the absence of a relationship between viral replication and target-organ injury. Instead, as shown for hepatitis B and C, in hepatitis A and E, experimental evidence supports a cytolytic CD8+ T-cell response as the instrument

of liver cell injury, in concert with or dwarfed by CD4+ helper T cells or CD4+ interferon γ -secreting cells. HEV has also been shown to interfere with host antiviral defenses, such as interferon signaling and effector function, and to downregulate interferon-stimulated genes. The demonstration of an activated innate immune response in patients with these hepatitides argues for a multitude of immunologic mechanisms in the pathogenesis of the acute liver injury resulting from HAV and HEV infection.

■ EXTRAHEPATIC MANIFESTATIONS

Immune complex-mediated tissue damage appears to play a pathogenic role in the extrahepatic manifestations of acute hepatitis B. The occasional prodromal serum sickness-like syndrome observed in acute hepatitis B appears to be related to the deposition in tissue blood vessel walls of HBsAg–anti-HBs circulating immune complexes, leading to activation of the complement system and depressed serum complement levels.

In patients with chronic hepatitis B, other types of immune-complex disease may be seen. Glomerulonephritis with the nephrotic syndrome is observed occasionally; HBsAg, immunoglobulin, and C3 deposition has been found in the glomerular basement membrane. Whereas generalized vasculitis (polyarteritis nodosa) develops in considerably <1% of patients with chronic HBV infection, 20–30% of patients with polyarteritis nodosa have HBsAg in serum (Chap. 363). In these patients, the affected small- and medium-size arterioles contain HBsAg, immunoglobulins, and complement components. Another extrahepatic manifestation of viral hepatitis, essential mixed cryoglobulinemia (EMC), was reported initially to be associated with hepatitis B. The disorder is characterized clinically by arthritis, cutaneous vasculitis (palpable purpura), and, occasionally, glomerulonephritis and serologically by the presence of circulating cryoprecipitable immune complexes of more than one immunoglobulin class (Chaps. 314 and 363). Many patients with this syndrome have chronic liver disease, but the association with HBV infection is limited; instead, a substantial proportion has chronic HCV infection, with circulating immune complexes containing HCV RNA. Immune-complex glomerulonephritis is another recognized extrahepatic manifestation of chronic hepatitis C (see “Complications and Sequelae,” below). Immune-complex disorders have been linked, albeit rarely, with both hepatitis A and E. In hepatitis E, rare neurologic (including Guillain-Barré syndrome), renal, pancreatic, and hematologic complications have been postulated to result from both immunologic mechanisms and/or direct extrahepatic-site infection with the virus.

■ PATHOLOGY

The typical morphologic lesions of all types of viral hepatitis are similar and consist of panlobular infiltration with mononuclear cells, hepatic cell necrosis, hyperplasia of Kupffer cells, and variable degrees of cholestasis. Hepatic cell regeneration is present, as evidenced by numerous mitotic figures, multinucleated cells, and “rosette” or “pseudoacinar” formation. The mononuclear infiltration consists primarily of small lymphocytes, although plasma cells and eosinophils occasionally are present. Liver cell damage consists of hepatic cell degeneration and necrosis, cell dropout, ballooning of cells, and acidophilic degeneration of hepatocytes (forming so-called Councilman or apoptotic bodies). Large hepatocytes with a ground-glass appearance of the cytoplasm may be seen in chronic but not in acute HBV infection; these cells contain HBsAg and can be identified histochemically with orcein or aldehyde fuchsin. In uncomplicated viral hepatitis, the reticulin framework is preserved.

In hepatitis C, the histologic lesion is often remarkable for a relative paucity of inflammation, a marked increase in activation of sinusoidal lining cells, lymphoid aggregates, the presence of fat (more frequent in genotype 3 and linked to increased fibrosis), and, occasionally, bile duct lesions in which biliary epithelial cells appear to be piled up without interruption of the basement membrane. Occasionally, microvesicular steatosis occurs in hepatitis D. In hepatitis E, a common histologic feature is marked cholestasis. A cholestatic variant of slowly resolving acute hepatitis A also has been described.

A more severe histologic lesion, *bridging hepatic necrosis*, also termed *subacute* or *confluent necrosis* or *interface hepatitis*, is observed occasionally in acute hepatitis. “Bridging” between lobules results from large areas of hepatic cell dropout, with collapse of the reticulin framework. Characteristically, the bridge consists of condensed reticulum, inflammatory debris, and degenerating liver cells that span adjacent portal areas, portal to central veins, or central vein to central vein. This lesion had been thought to have prognostic significance; in many of the originally described patients with this lesion, a subacute course terminated in death within several weeks to months, or severe chronic hepatitis and cirrhosis developed; however, the association between bridging necrosis and a poor prognosis in patients with acute hepatitis has not been upheld. Therefore, although demonstration of this lesion in patients with chronic hepatitis has prognostic significance (Chap. 341), its demonstration during acute hepatitis is less meaningful, and liver biopsies to identify this lesion are no longer undertaken routinely in patients with acute hepatitis. In *massive hepatic necrosis* (fulminant hepatitis, “acute yellow atrophy”), the striking feature at postmortem examination is the finding of a small, shrunken, soft liver. Histologic examination reveals massive necrosis and dropout of liver cells of most lobules with extensive collapse and condensation of the reticulin framework. When histologic documentation is required in the management of fulminant or very severe hepatitis, a biopsy can be done by the angiographically guided transjugular route, which permits the performance of this invasive procedure in the presence of severe coagulopathy.

Immunohistochemical and electron-microscopic studies have localized HBsAg to the cytoplasm and plasma membrane of infected liver cells. In contrast, HBcAg predominates in the nucleus, but, occasionally, scant amounts are also seen in the cytoplasm and on the cell membrane. HDV antigen is localized to the hepatocyte nucleus, whereas HAV and HCV antigens are localized to the cytoplasm. Hepatitis E ORF-2 protein staining is distributed in both a cytoplasmic and nuclear pattern.

■ EPIDEMIOLOGY AND GLOBAL FEATURES

Before the availability of serologic tests for hepatitis viruses, all viral hepatitis cases were labeled either as “infectious” or “serum” hepatitis. Modes of transmission overlap, however, and *a clear distinction among the different types of viral hepatitis cannot be made solely based on clinical or epidemiologic features* (Table 339-2). The most accurate means to distinguish the various types of viral hepatitis involves specific serologic testing.

Hepatitis A *This agent is transmitted almost exclusively by the fecal-oral route.* Person-to-person spread of HAV is enhanced by poor personal hygiene and overcrowding; large outbreaks as well as sporadic cases have been traced to contaminated food, water, milk, frozen raspberries and strawberries, green onions imported from Mexico, and shellfish (e.g., scallops imported from the Philippines used to make sushi, the culprit identified in a 2016 Hawaiian outbreak). Intrafamily and intrainstitutional spreads are also common. Early epidemiologic observations supported a predilection for hepatitis A to occur in late fall and early winter. In temperate zones, epidemic waves have been recorded every 5–20 years as new segments of nonimmune population appeared; however, in developed countries, the incidence of hepatitis A has been declining, presumably as a function of improved sanitation, and these cyclic patterns are no longer observed. No HAV carrier state has been identified after acute hepatitis A; perpetuation of the virus in nature depends presumably on nonepidemic, inapparent subclinical infection, ingestion of contaminated food or water in, or imported from, endemic areas, and/or contamination linked to environmental reservoirs.

In the general population, anti-HAV, a marker for previous HAV infection, increases in prevalence as a function of increasing age and of decreasing socioeconomic status. In the 1970s, serologic evidence of prior hepatitis A infection occurred in ~40% of urban populations in the United States, most of whose members never recalled having had a symptomatic case of hepatitis. In subsequent decades, however, the

TABLE 339-2 Clinical and Epidemiologic Features of Viral Hepatitis

FEATURE	HAV	HBV	HCV	HDV	HEV
Incubation (days)	15–45, mean 30	30–180, mean 60–90	15–160, mean 50	30–180, mean 60–90	14–60, mean 40
Onset	Acute	Insidious or acute	Insidious or acute	Insidious or acute	Acute
Age preference	Children, young adults	Young adults (sexual and percutaneous), babies, toddlers	Any age, but more common in adults	Any age (similar to HBV)	Epidemic cases: young adults (20–40 years); sporadic cases: older adults (>60)
Transmission					
Fecal-oral	+++	–	–	–	+++
Percutaneous	Unusual	+++	+++	+++	–
Perinatal	–	+++	‡ ^a	+	–
Sexual	±	++	‡ ^a	++	–
Clinical					
Severity	Mild	Occasionally severe	Moderate	Occasionally severe	Mild
Fulminant	0.1%	0.1–1%	0.1%	5–20% ^b	1–2% ^c
Progression to chronicity	None	Occasional (1–10%)	Common (85%)	Common ^d	None ^e
Carrier	None	(90% of neonates)	1.5–3.2%	Variable ^g	None
Cancer	None	0.1–30% ^f + (neonatal infection)	+	±	None
Prognosis	Excellent	Worse with age, debility	Moderate	Acute, good; chronic, poor	Good
Prophylaxis	Ig, inactivated vaccine	HBIG, recombinant vaccine	None	HBV vaccine (none for HBV carriers)	Vaccine
Therapy	None	Interferon ^h Lamivudine ^h Adefovir ^h Pegylated interferon ⁱ Entecavir ^j Telbivudine ^j Tenofovir disoproxil fumarate ^j Tenofovir alafenamide ^j	Pegylated interferon ribavirin, ^h telaprevir, ^h boceprevir, ^h simeprevir, ^h sofosbuvir, ^h ledipasvir, paritaprevir/ ritonavir, ^h ombitasvir, ^h dasabuvir, ^h daclatasvir, ^h velpatasvir, grazoprevir, elbasvir, glecaprevir, pibrentasvir, voxilaprevir	Pegylated interferon ±	None ^l

^aPrimarily with HIV co-infection and high-level viremia in index case; more likely in persons with multiple sex partners or sexually transmitted diseases; risk ~5%. ^bUp to 5% in acute HBV/HDV co-infection; up to 20% in HDV superinfection of chronic HBV infection. ^c10–20% in pregnant women. ^dIn acute HBV/HDV co-infection, the frequency of chronicity is the same as that for HBV; in HDV superinfection, chronicity is invariable. ^eExcept as observed in immunosuppressed liver allograft recipients or other immunosuppressed hosts. ^fVaries considerably throughout the world and in subpopulations within countries; see text. ^gCommon in Mediterranean countries; rare in North America and western Europe. ^hNo longer recommended or not included in first-line therapy. ⁱFirst-line agents. ^jAnecdotal reports and retrospective studies suggest that pegylated interferon and/or ribavirin are effective in treating chronic hepatitis E, observed in immunocompromised persons; ribavirin monotherapy has been used successfully in acute, severe hepatitis E.

Abbreviation: HBIG, hepatitis B immunoglobulin. See text for other abbreviations.

prevalence of anti-HAV declined in the United States. In developing countries, exposure, infection, and subsequent immunity are almost universal in childhood. As the frequency of subclinical childhood infections declines in developed countries, a susceptible cohort of adults emerges. Hepatitis A tends to be more symptomatic in adults; therefore, paradoxically, as the frequency of HAV infection declines, the likelihood of clinically apparent, even severe, HAV illnesses increases in the susceptible adult population. Travel to endemic areas is a common source of infection for adults from nonendemic areas. Important recognized epidemiologic foci of HAV infection include childcare centers, neonatal intensive care units, promiscuous men who have sex with men, injection drug users, and unvaccinated close contacts of newly arrived international adopted children, most of whom emanate from countries with intermediate-to-high hepatitis A endemicity. Although hepatitis A is rarely bloodborne, several outbreaks have been recognized in recipients of clotting-factor concentrates. In the United States, the introduction of hepatitis A vaccination programs among children from high-incidence states has resulted in a >70% reduction in the annual incidence of new HAV infections and has shifted the burden of new infections from children to adults. In the 2007–2012 U.S. Public Health Service National Health and Nutrition Examination Survey (NHANES), the prevalence of anti-HAV in the U.S. population aged ≥20 years had declined to 24.2% from the 29.5% measured in NHANES 1999–2006. While universal childhood vaccination accounted for a high prevalence of vaccine-induced immunity in children aged 2–19 years, the lowest age-specific prevalence of anti-HAV (16.1–17.6%) occurred in adults in the fourth and fifth decades (aged 30–49 years). This is a subgroup of the population who remain

susceptible to acute hepatitis A acquired during travel to endemic areas and from contaminated foods, especially those imported from endemic countries. Recognized initially in San Diego, California, in 2016, widespread person-to-person outbreaks, attributed to fecally contaminated environments, of acute hepatitis A occurred primarily among homeless persons and persons who were using injection drugs. Ultimately, this outbreak extended to at least 32 states (highest number of cases in Kentucky), and by March 2020, 31,950 cases were reported, resulting in 19,548 hospitalizations (61% of cases) and 322 deaths (1% of reported cases, 1.6% of hospitalized cases). The increased clinical severity, rate of hospitalization, and death in these outbreaks can be attributed to their involving an older population (mean age ranging from 36 to 42 years), born before the introduction of universal childhood hepatitis A vaccination and in whom clinical severity, as noted above, is higher than in children. Moreover, the affected homeless and drug-using populations suffer from multiple comorbidities (including HBV or HCV co-infection) and disparities in access to health care. Addressing this multistate outbreak has required a vigorous hepatitis A vaccination effort as well as environmental sanitation/hygiene and education among these susceptible populations.

Hepatitis B Percutaneous inoculation has long been recognized as a major route of hepatitis B transmission, but the outmoded designation “serum hepatitis” is an inaccurate label for the epidemiologic spectrum of HBV infection. As detailed below, most of the hepatitis transmitted by blood transfusion is not caused by HBV; moreover, in approximately two-thirds of patients with acute type B hepatitis, no history of an identifiable percutaneous exposure can be elicited. We

now recognize that many cases of hepatitis B result from less obvious modes of nonpercutaneous or covert percutaneous transmission. HBsAg has been identified in almost every body fluid from infected persons, and at least some of these body fluids—most notably semen and saliva—are infectious, albeit less so than serum, when administered percutaneously or nonpercutaneously to experimental animals. Among the nonpercutaneous modes of HBV transmission, oral ingestion has been documented as a potential but inefficient route of exposure. By contrast, the two nonpercutaneous routes considered to have the greatest impact are intimate (especially sexual) contact and perinatal transmission.

In sub-Saharan Africa, intimate contact among toddlers is considered instrumental in contributing to the maintenance of the high frequency of hepatitis B in the population. Perinatal transmission occurs primarily in infants born to mothers with chronic hepatitis B or (rarely) mothers with acute hepatitis B during the third trimester of pregnancy or during the early postpartum period. Perinatal transmission is uncommon in North America and western Europe but occurs with great frequency and is the most important mode of HBV perpetuation in East Asia and developing countries. Although the precise mode of perinatal transmission is unknown, and although ~10% of infections may be acquired in utero, epidemiologic evidence suggests that most infections occur approximately at the time of delivery and are not related to breast-feeding (which is not contraindicated in women with hepatitis B). The likelihood of perinatal transmission of HBV correlates with the presence of HBeAg and high-level viral replication; 90% of HBeAg-positive mothers but only 10–15% of anti-HBe-positive mothers transmit HBV infection to their offspring. In most cases, acute infection in the neonate is clinically asymptomatic, but the child is very likely to remain chronically infected.

The 250–290 million persons with chronic HBV infection in the world constitute the main reservoir of hepatitis B in human beings. Whereas serum HBsAg is infrequent (0.1–0.5%) in normal populations in the United States and western Europe, a prevalence of up to 5–10% has been found in East Asia, sub-Saharan Africa, and tropical countries; the prevalence can be even higher in certain high-risk groups, including persons with Down's syndrome, lepromatous leprosy, leukemia, Hodgkin's disease, polyarteritis nodosa, and chronic renal disease on hemodialysis, as well as in injection drug users.

Other groups with high rates of HBV infection include spouses of acutely infected persons; sexually promiscuous persons (especially promiscuous men who have sex with men); health care workers exposed to blood; persons who require repeated transfusions especially with pooled blood-product concentrates (e.g., hemophiliacs); residents and staff of custodial institutions for the developmentally handicapped; prisoners; and, to a lesser extent, family members of chronically infected patients. In volunteer blood donors, the prevalence of anti-HBs, a reflection of previous HBV infection, ranges from 5 to 10%, but the prevalence is higher in lower socioeconomic strata, older age groups, and persons—including those mentioned above—exposed to blood products. Because of highly sensitive virologic screening (antigen, antibody, and nucleic acid testing) of donor blood, the risk of acquiring HBV infection from a blood transfusion is 1 in 230,000 to 1 in 346,000.

Prevalence of infection, modes of transmission, and human behavior conspire to mold geographically different epidemiologic patterns of HBV infection. In East Asia and Africa, hepatitis B, a disease of the newborn and young children, is perpetuated by a cycle of maternal-neonatal spread. In North America and western Europe, hepatitis B is primarily a disease of adolescence and early adulthood, the time of life when intimate sexual contact and recreational and occupational percutaneous exposures tend to occur. To some degree, however, this dichotomy between high-prevalence and low-prevalence geographic regions has been minimized by immigration from high-prevalence to low-prevalence areas. For example, in the United States, NHANES data from 2007 to 2012 revealed an overall prevalence of current HBV infection (detectable HBsAg) of 0.3%; however, the prevalence in Asian persons, 93% of whom were foreign-born, was tenfold higher, 3.1%, representing 50% of the U.S. national disease burden. As a result of

adoption of safe behaviors in high-risk groups as well as screening and vaccination programs, the incidence of newly reported HBV infections fell by >80% in the United States during the 1990s (with a low of 3050 reported cases in 2013). Paralleling that trend, the imbalance between cases in U.S.-born and foreign-born persons widened; currently, imported cases in non-U.S.-born persons outnumber domestic cases by manyfold; in NHANES 1999–2016, the 2016 prevalence of HBV infection was 0.24% in foreign-born versus 0.06% in U.S.-born persons; in Asian persons, the 2016 prevalence of HBV infections was 3.85% in foreign-born versus 0.79% in U.S.-born persons. The introduction of hepatitis B vaccine in the early 1980s and adoption of universal childhood vaccination policies in many countries resulted in a dramatic, ~90% decline in the incidence of new HBV infections in those countries as well as in the dire consequences of chronic infection, including hepatocellular carcinoma. In the United States, as demonstrated in NHANES 2007–2012, following the 1991 implementation of universal childhood vaccination, HBsAg seropositivity had declined in children aged 6–19 years to as low as 0.03%, an ~85% reduction. Populations and groups for whom HBV infection screening is recommended are listed in Table 339-3.

Hepatitis D Infection with HDV has a worldwide distribution, but two epidemiologic patterns exist. In Mediterranean countries (northern Africa, southern Europe, the Middle East), HDV infection is endemic among those with hepatitis B, and the disease is transmitted predominantly by nonpercutaneous means, especially close personal contact. In nonendemic areas, such as the United States (where hepatitis D is rare among persons with chronic hepatitis B) and northern Europe, HDV infection is confined to persons exposed frequently to blood and blood products, primarily injection drug users (especially in HIV-infected injection drug users) and hemophiliacs. In the United States, the prevalence of HDV infection in the national population was 0.02% in NHANES 1999–2012 and 0.11% in NHANES 2011–2016; however, among HBsAg-positive persons, the prevalence of HDV infection is highest in injection drug users (11–36%) and hemophiliacs (19%). HDV infection can be introduced into a population through drug users or by migration of persons from endemic to nonendemic areas. Thus, patterns of population migration and human behavior facilitating percutaneous contact play important roles in the introduction and amplification of HDV infection. Occasionally, the migrating epidemiology of hepatitis D is expressed in explosive outbreaks of severe hepatitis, such as those that have occurred in remote South American villages (e.g., "Lábrea fever" in the Amazon basin) as well as in urban centers in the United States. Ultimately, such outbreaks

TABLE 339-3 High-Risk Populations for Whom HBV Infection Screening Is Recommended

Persons born in countries/regions with a high ($\geq 8\%$) and intermediate ($\geq 2\%$) prevalence of HBV infection including immigrants and adopted children and including persons born in the United States who were not vaccinated as infants and whose parents emigrated from areas of high HBV endemicity
Household and sexual contacts of persons with hepatitis B
Babies born to HBsAg-positive mothers
Persons who have used injection drugs
Persons with multiple sexual contacts or a history of sexually transmitted disease
Men who have sex with men
Inmates of correctional facilities
Persons with elevated alanine or aspartate aminotransferase levels
Blood/plasma/organ/tissue/semen donors
Persons with HCV or HIV infection
Hemodialysis patients
Pregnant women
Persons who are the source of blood or body fluids that would be an indication for postexposure prophylaxis (e.g., needlestick, mucosal exposure, sexual assault)
Persons who require immunosuppressive or cytotoxic therapy (including anti-tumor necrosis factor α therapy for rheumatologic or inflammatory bowel disorders)

of hepatitis D—either of co-infections with acute hepatitis B or of superinfections in those already infected with HBV—may blur the distinctions between endemic and nonendemic areas. On a global scale, HDV infection declined at the end of the 1990s. Even in Italy, an HDV-endemic area, public health measures introduced to control HBV infection (e.g., mass hepatitis B vaccination) resulted during the 1990s in a 1.5%/year reduction in the prevalence of HDV infection. Still, the frequency of HDV infection during the first decade of the twenty-first century has not fallen below levels reached during the 1990s; the reservoir has been sustained by survivors infected during 1970–1980 and recent immigrants from still-endemic (e.g., eastern Europe and Central Asia) to less-endemic countries. The current global prevalence of HDV infection has been estimated at 62–72 million people. Of the eight HDV genotypes, genotype 1 is distributed worldwide, while the others are more geographically confined (e.g., genotypes 2 and 4 in the Far East, 3 in South America, and 5–8 in Africa).

Hepatitis C Routine screening of blood donors for HBsAg and the elimination of commercial blood sources in the early 1970s reduced the frequency of, but did not eliminate, transfusion-associated hepatitis. During the 1970s, the likelihood of acquiring hepatitis after transfusion of voluntarily donated, HBsAg-screened blood was ~10% per patient (up to 0.9% per unit transfused); 90–95% of these cases were classified, based on serologic exclusion of hepatitis A and B, as “non-A, non-B” hepatitis. For patients requiring transfusion of pooled products, such as clotting factor concentrates, the risk was even higher, up to 20–30%.

During the 1980s, voluntary self-exclusion of blood donors with risk factors for AIDS and then the introduction of donor screening for anti-HIV reduced further the likelihood of transfusion-associated hepatitis to <5%. During the late 1980s and early 1990s, the introduction first of “surrogate” screening tests for non-A, non-B hepatitis (alanine aminotransferase [ALT] and anti-HBc, both shown to identify blood donors with a higher likelihood of transmitting non-A, non-B hepatitis to recipients) and, subsequently, after the discovery of HCV, progressively more sensitive immunoassays for anti-HCV and then the application of automated PCR testing of donated blood for HCV RNA reduced the risk of transfusion-associated hepatitis C even further, to almost imperceptible levels ranging between 1 in 2.3 million transfusions to 1 in 4.7 million transfusions.

In addition to being transmitted by transfusion, hepatitis C can be transmitted by other percutaneous routes, such as injection drug use. This virus can be transmitted by occupational exposure to blood, and the likelihood of infection is increased in hemodialysis units. Although the frequency of transfusion-associated hepatitis C fell as a result of blood-donor screening, the *overall* frequency of reported hepatitis C cases did not change until the 1990s, when the overall frequency of reported cases fell by 80%, in parallel with a reduction in the number of new cases in injection drug users, the source of most of the HCV reservoir. After the exclusion of anti-HCV-positive plasma units from the donor pool, rare, sporadic instances occurred of hepatitis C among recipients of immunoglobulin preparations for intravenous (but not intramuscular) use.

Serologic evidence for HCV infection occurs in 90% of patients with a history of transfusion-associated hepatitis (almost all occurring before 1992, when second-generation HCV screening tests were introduced); hemophiliacs and others treated with clotting factors; injection drug users; 60–70% of patients with sporadic “non-A, non-B” hepatitis who lack identifiable risk factors; 0.5% of volunteer blood donors; and, in the NHANES survey conducted in the United States between 1999 and 2002, 1.6% of the general population in the United States, which translated into 4.1 million persons (3.2 million with viremia), the majority of whom were unaware of their infections. Moreover, such population surveys do not include higher-risk groups such as incarcerated persons, homeless persons, and active injection drug users, indicating that the actual prevalence is even higher (estimated to add an additional 1 million with anti-HCV antibody and 0.8 million with HCV RNA in a later cohort assessed in 2003–2010). Comparable frequencies of HCV infection occur in most countries around the world, with 71 million persons infected worldwide, but extraordinarily

high prevalences of HCV infection occur in certain countries such as Egypt, where >20% of the population (as high as 50% in persons born prior to 1960) in some cities is infected. The high frequency in Egypt is attributable to contaminated equipment used for medical procedures and unsafe injection practices in the 1950s to 1980s (during a campaign to eradicate schistosomiasis with intravenous tartar emetic). Thanks to a 2018–2019 Egyptian government program to screen its entire adult population (79% participation among >60 million people) for hepatitis C and treat infected persons (2.2 million, 4.6% of those screened; of the 83% with a documented outcome, 99% were cured; the cost to identify and cure a person was \$130) with generic versions of direct-acting antiviral (DAA) therapy (Chap. 341), hepatitis C has been nearly eliminated there.

In the United States, African Americans and Mexican Americans have higher frequencies of HCV infection than whites. Data from NHANES showed that between 1988 and 1994, 30- to 40-year-old men had the highest prevalence of HCV infection; however, in the NHANES survey conducted between 1999 and 2002, the peak age decile had shifted to those aged 40–49 years; an increase in hepatitis C-related mortality has paralleled this secular trend, increasing since 1995 predominantly in the 45- to 65-year age group. Thus, despite an 80% reduction in new reported HCV infections during the 1990s, the prevalence of HCV infection in the population was sustained by an aging cohort that had acquired their infections three to four decades earlier, during the 1960s and 1970s, as a result predominantly of self-inoculation with recreational drugs. Retrospective phylogenetic mapping of >45,000 HCV genotype 1a isolates revealed that the hepatitis C epidemic emerged in the United States between 1940 and 1965, peaking in 1950 and aligning temporally with the post–World War II expansion of medical procedures (including reuse of glass syringes). Thus, HCV was amplified iatrogenically not only in Egypt but also in the United States; in the United States, the seeds sown by medical procedures in the 1950s were reaped in the 1960s and 1970s among transfusion recipients and injection drug users, even those whose drug use was confined to brief adolescent experimentation.

In NHANES 2003–2010, the prevalence of HCV infection (HCV RNA reactivity) in the United States had actually fallen to 1% (2.7 million persons) from 1.3% (3.2 million) the decade before (NHANES 1999–2002), attributable to deaths among the HCV-infected population. In NHANES data from 2010–2014, the prevalence of current HCV infection (HCV RNA reactivity) had fallen even lower, to 0.65% (1.7 million persons), coinciding with and attributable to the introduction of highly effective, oral DAA drugs (Chap. 341). As deaths resulting from HIV infection fell after 1999, age-adjusted mortality associated with HCV infection surpassed that of HIV infection in 2007; >70% of HCV-associated deaths occurred in the “baby boomer” cohort born between 1945 and 1965. By 2012, HCV mortality had surpassed deaths from HIV, tuberculosis, hepatitis B, and 57 other notifiable infectious diseases (i.e., *all* infectious diseases) reported to the CDC. In NHANES 1999–2002, compared to the 1.6% prevalence of HCV infection in the population at large, the prevalence in the 1945–1965 birth cohort was 3.2%, representing three-quarters of all infected persons. Therefore, in 2012, the CDC and, in 2013, the U.S. Preventive Services Task Force (USPSTF) recommended that all persons born between 1945 and 1965 be screened for hepatitis C, without ascertainment of risk, a recommendation shown to be cost-effective and predicted to identify 800,000 infected persons. Because of the availability of highly effective antiviral therapy, such screening would have the potential to avert 200,000 cases of cirrhosis and 47,000 cases of hepatocellular carcinoma and to prevent 120,000 hepatitis-related deaths; with the availability of the new generation of DAAs (efficacy >95%, see Chap. 341), screening baby boomers and treating those with hepatitis C have been predicted to reduce the HCV-associated disease burden by 50–70% through 2050.

Still, persons with chronic hepatitis C identified by 1945–1965 birth-cohort screening are older than 50, and by the time they are identified, >20% already have advanced liver disease. In 2020, based on (1) the 95–99% efficacy of all-oral, well-tolerated, highly effective DAAs; (2) the demonstration that the endpoint of DAA therapy (sustained

2574 virologic response) was associated with a marked decrease in liver and all-cause mortality, cirrhosis, and hepatocellular carcinoma ([Chap. 341](#)); (3) a reduction in the initially high cost of DAA therapy; (4) the demonstration of higher cost-effectiveness of screening all adults rather than birth-cohort screening; and (5) the shifting demographics of HCV infection (see below), especially since 2010, toward a younger population exposed through injection drug use, the American Association for the Study of Liver Diseases and the Infectious Diseases Society of America as well as the USPSTF and CDC expanded recommended hepatitis C screening to all adolescents and adults aged 18–79 (and because of the substantial increase in HCV infections among women of child-bearing age [age 20–39], expanded such screening to pregnant women).

Hepatitis C accounts for 40% of chronic liver disease and, before the introduction of high-efficacy DAA therapy, was the most frequent indication for liver transplantation; hepatitis C is estimated to account for 8000–10,000 deaths per year in the United States. The distribution of HCV genotypes varies in different parts of the world. Worldwide, genotype 1 is the most common. In the United States, genotype 1 accounts for 70% of HCV infections, whereas genotypes 2 and 3 account for the remaining 30%; among African Americans, the frequency of genotype 1 is even higher (i.e., 90%). Genotype 4 predominates in Egypt; genotype 5 is localized to South Africa, genotype 6 to Hong Kong, and genotype 7 to Central Africa. Most asymptomatic blood donors found to have anti-HCV and ~20–30% of persons with reported cases of acute hepatitis C do not fall into a recognized risk group; however, many such blood donors do recall risk-associated behaviors when questioned carefully.

As a bloodborne infection, HCV potentially can be transmitted sexually and perinatally; however, both modes of transmission are inefficient for hepatitis C. Although 10–15% of patients with acute hepatitis C report having potential sexual sources of infection, most studies have failed to identify sexual transmission of this agent. The chances of sexual and perinatal transmission have been estimated to be ~5% but have shown in a prospective study to be only 1% between monogamous sexual partners, well below comparable rates for HIV and HBV infections. Moreover, sexual transmission appears to be confined to such subgroups as persons with multiple sexual partners and sexually transmitted diseases; for example, isolated clusters of sexually transmitted HCV infection have been reported in HIV-infected men who have sex with men. Breast-feeding does not increase the risk of HCV infection between an infected mother and her infant. Infection of health workers is not dramatically higher than among the general population; however, health workers are more likely to acquire HCV infection through accidental needle punctures, the efficiency of which is ~3%. Infection of household contacts is rare as well.

Besides persons born between 1945 and 1965, other groups with an increased frequency of HCV infection are listed in [Table 339-4](#). In immunosuppressed individuals, levels of anti-HCV may be undetectable, and a diagnosis may require testing for HCV RNA. Although new acute cases of hepatitis C are rare outside of the injection drug-using community, newly diagnosed cases are common among otherwise healthy persons who experimented briefly with injection drugs, as noted above, four or five decades earlier. Such instances usually remain unrecognized for years, until unearthed by laboratory screening for routine medical examinations, insurance applications, and attempted blood donation. Although, overall, the annual incidence of new HCV infections has continued to fall, the rate of new infections has been increasing since 2002, has accelerated since 2010 (tripling from 0.3/100,000 to 1.2/100,000 between 2009 and 2018), and has been amplified by the recent epidemic of opioid use in a new cohort of young injection drug users aged 20–39 years (accounting for a 3.8-fold increase in cases between 2010 and 2017 and for more than two-thirds of all acute cases), who, unlike older cohorts, had not learned to take precautions to prevent bloodborne infections. Reflecting this emerging development, the prevalence of current HCV infection (HCV RNA reactivity) in the United States rose from 0.65% (1.7 million persons) in a 2010–2014 NHANES analysis to 0.84% (2.04 million persons) in a 2013–2014 NHANES analysis. Moreover, based on an estimate

TABLE 339-4 High-Risk Populations for Whom HCV-Infection Screening Is Recommended

All adults aged 18–79 should be screened, a recommendation that supplants the earlier focus on persons born between 1945 and 1965
Persons who have ever used injection drugs
Persons with HIV infection
Hemophiliacs treated with clotting factor concentrates prior to 1987
Persons who have ever undergone long-term hemodialysis
Persons with unexplained elevations of aminotransferase levels
Transfusion or transplantation recipients prior to July 1992
Recipients of blood or organs from a donor found to be positive for hepatitis C
Children born to women with hepatitis C
Health care, public safety, and emergency medical personnel following needle injury or mucosal exposure to HCV-contaminated blood
Sexual partners of persons with hepatitis C infection
Pregnant women

of populations excluded from this NHANES analysis, the prevalence would be even higher, 0.93% (2.27 million persons). This late temporal trend was attributed to the increase of acute cases in injections drug users, driven by increases in states most affected by the opioid/injection drug use epidemic. Also, in parallel with this trend, the prevalence of HCV infection in women aged 15–44 years (of child-bearing age) doubled between 2016 and 2014; accordingly, screening of pregnant women for HCV infection is now recommended as well.

Hepatitis E This type of hepatitis, identified in India, Asia, Africa, the Middle East, and Central America (endemic areas), resembles hepatitis A in its primarily enteric mode of spread. The commonly recognized cases occur after contamination of water supplies such as after monsoon flooding, but sporadic, isolated cases occur. An epidemiologic feature that distinguishes HEV from other enteric agents is the rarity of secondary person-to-person spread from infected persons to their close contacts. Large waterborne outbreaks in endemic areas are linked to genotypes 1 and 2, arise in populations that are immune to HAV, favor young adults, and account for antibody prevalences of 30–80%. The worldwide annual incidence of acute HEV infections has been estimated conservatively to be at least 20 million (of which 3.3 million are symptomatic), rendering HEV infection as the most common cause of acute viral hepatitis. In nonendemic areas of the world, such as the United States, clinically apparent acute hepatitis E is extremely rare; however, during the 1988–1994 NHANES survey conducted by the U.S. Public Health Service, the prevalence of anti-HEV was 21%, reflecting subclinical infections, infection with genotypes 3 and 4, predominantly in older males (>60 years). A repeat NHANES study in 2009–2010, however, showed a substantial 70% two-decade reduction in anti-HEV to only 6%, more consistent with the rarity of acute hepatitis E in the United States than the previous NHANES result would suggest and perhaps a reflection of a more specific anti-HEV assay used in the second time period. Again, older age was associated with anti-HEV seropositivity. In nonendemic areas, HEV accounts for only a small proportion of cases of sporadic (labeled “autochthonous” or indigenous) hepatitis; however, cases imported from endemic areas have been found in the United States. Evidence supports a zoonotic reservoir for HEV primarily in swine (but also in deer, camels, and rabbits), which may account for the mostly subclinical infections primarily of genotypes 3 and 4 in nonendemic areas. A previously unrecognized high distribution of HEV infection, linked to uncooked or undercooked pork-product ingestion, has been discovered in western Europe (e.g., in Germany, an estimated annual incidence of 300,000 cases and a 17% prevalence of anti-HEV among adults; in France, a 22% prevalence of anti-HEV in healthy blood donors).

CLINICAL AND LABORATORY FEATURES

Symptoms and Signs Acute viral hepatitis occurs after an incubation period that varies according to the responsible agent. Generally, incubation periods for hepatitis A range from 15 to 45 days (mean,

4 weeks), for hepatitis B and D from 30 to 180 days (mean, 8–12 weeks), for hepatitis C from 15 to 160 days (mean, 7 weeks), and for hepatitis E from 14 to 60 days (mean, 5–6 weeks). The *prodromal symptoms* of acute viral hepatitis are systemic and quite variable. Constitutional symptoms of anorexia, nausea and vomiting, fatigue, malaise, arthralgias, myalgias, headache, photophobia, pharyngitis, cough, and coryza may precede the onset of jaundice by 1–2 weeks. The nausea, vomiting, and anorexia are frequently associated with alterations in olfaction and taste. A low-grade fever between 38° and 39°C (100°–102°F) is more often present in hepatitis A and E than in hepatitis B or C, except when hepatitis B is heralded by a serum sickness-like syndrome; rarely, a fever of 39.5°–40°C (103°–104°F) may accompany the constitutional symptoms. Dark urine and clay-colored stools may be noticed by the patient from 1–5 days before the onset of clinical jaundice.

With the onset of *clinical jaundice*, the constitutional prodromal symptoms usually diminish, but in some patients, mild weight loss (2.5–5 kg) is common and may continue during the entire icteric phase. The liver becomes enlarged and tender and may be associated with right upper quadrant pain and discomfort. Infrequently, patients present with a cholestatic picture, suggesting extrahepatic biliary obstruction. Splenomegaly and cervical adenopathy are present in 10–20% of patients with acute hepatitis. Rarely, a few spider angiomas appear during the icteric phase and disappear during convalescence. During the *recovery phase*, constitutional symptoms disappear, but usually some liver enlargement and abnormalities in liver biochemical tests are still evident. The duration of the posticteric phase is variable, ranging from 2 to 12 weeks, and is usually more prolonged in acute hepatitis B and C. Complete clinical and biochemical recovery is to be expected 1–2 months after all cases of hepatitis A and E and 3–4 months after the onset of jaundice in three-quarters of uncomplicated, self-limited cases of hepatitis B and C (among healthy adults, acute hepatitis B is self-limited in 95–99%, whereas hepatitis C is self-limited in only ~15–20%). In the remainder, biochemical recovery may be delayed. A substantial proportion of patients with viral hepatitis never become icteric.

Infection with HDV can occur in the presence of acute or chronic HBV infection; the duration of HBV infection determines the duration of HDV infection. When acute HDV and HBV infections occur simultaneously, clinical and biochemical features may be indistinguishable from those of HBV infection alone, although occasionally, they are more severe. As opposed to patients with *acute* HBV infection, patients with *chronic* HBV infection can support HDV replication indefinitely, as when acute HDV infection occurs in the presence of a nonresolving acute HBV infection or, more commonly, when acute hepatitis D is superimposed on underlying chronic hepatitis B. In such cases, the HDV superinfection appears as a clinical exacerbation or an episode resembling acute viral hepatitis in someone already chronically infected with HBV. Superinfection with HDV in a patient with chronic hepatitis B often leads to clinical deterioration (see below).

In addition to superinfections with other hepatitis agents, acute hepatitis-like clinical events in persons with chronic hepatitis B may accompany spontaneous HBeAg to anti-HBe seroconversion or spontaneous reactivation (i.e., reversion from relatively nonreplicative to replicative infection). Such reactivations can occur as well in therapeutically immunosuppressed patients with chronic HBV infection when cytotoxic/immunosuppressive drugs are withdrawn; in these cases, restoration of immune competence is thought to allow resumption of previously checked cell-mediated immune cytolysis of HBV-infected hepatocytes. Occasionally, acute clinical exacerbations of chronic hepatitis B may represent the emergence of a precore mutant (see “*Virology and Etiology*”), and the subsequent course in such patients may be characterized by periodic exacerbations. Cytotoxic chemotherapy can lead to reactivation of chronic hepatitis C as well, and treatment with other immunomodulators, such as monoclonal antibodies against anti-TNF- α and other cytokines and especially the B-cell (CD20)-depleting antibody rituximab, can lead to reactivation of both hepatitis B and C.

Laboratory Features The serum aminotransferases aspartate aminotransferase (AST) and ALT (previously designated SGOT and

SGPT) increase to a variable degree during the prodromal phase of acute viral hepatitis and precede the rise in bilirubin level (Figs. 339-2 and 339-4). The level of these enzymes, however, does not correlate well with the degree of liver cell damage. Peak levels vary from ~400 to ~4000 IU or more; these levels are usually reached at the time the patient is clinically icteric and diminish progressively during the recovery phase of acute hepatitis. The diagnosis of anicteric hepatitis is based on clinical features and on aminotransferase elevations.

Jaundice is usually visible in the sclera or skin when the serum bilirubin value is >43 $\mu\text{mol/L}$ (2.5 mg/dL). When jaundice appears, the serum bilirubin typically rises to levels ranging from 85 to 340 $\mu\text{mol/L}$ (5–20 mg/dL). The serum bilirubin may continue to rise despite falling serum aminotransferase levels. In most instances, the total bilirubin is equally divided between the conjugated and unconjugated fractions. Bilirubin levels >340 $\mu\text{mol/L}$ (20 mg/dL) extending and persisting late into the course of viral hepatitis are more likely to be associated with severe disease. In certain patients with underlying hemolytic anemia, however, such as glucose-6-phosphate dehydrogenase deficiency and sickle cell anemia, a high serum bilirubin level is common, resulting from superimposed hemolysis. In such patients, bilirubin levels >513 $\mu\text{mol/L}$ (30 mg/dL) have been observed and are not necessarily associated with a poor prognosis.

Neutropenia and lymphopenia are transient and are followed by a relative lymphocytosis. Atypical lymphocytes (varying between 2 and 20%) are common during the acute phase. Measurement of the prothrombin time (PT) is important in patients with acute viral hepatitis, because a prolonged value may reflect a severe hepatic synthetic defect, signify extensive hepatocellular necrosis, and indicate a worse prognosis. Occasionally, a prolonged PT may occur with only mild increases in the serum bilirubin and aminotransferase levels. Prolonged nausea and vomiting, inadequate carbohydrate intake, and poor hepatic glycogen reserves may contribute to hypoglycemia noted occasionally in patients with severe viral hepatitis. Serum alkaline phosphatase may be normal or only mildly elevated, whereas a fall in serum albumin is uncommon in uncomplicated acute viral hepatitis. In some patients, mild and transient steatorrhea has been noted, as well as slight microscopic hematuria and minimal proteinuria.

A diffuse but mild elevation of the γ globulin fraction is common during acute viral hepatitis. Serum IgG and IgM levels are elevated in about one-third of patients during the acute phase of viral hepatitis, but the serum IgM level is elevated more characteristically during acute hepatitis A. During the acute phase of viral hepatitis, antibodies to smooth muscle and other cell constituents may be present, and low titers of rheumatoid factor, nuclear antibody, and heterophile antibody can also be found occasionally. In hepatitis C and D, antibodies to LKM may occur; however, the species of LKM antibodies in the two types of hepatitis are different from each other as well as from the LKM antibody species characteristic of autoimmune hepatitis type 2 (Chap. 341). The autoantibodies in viral hepatitis are nonspecific and can also be associated with other viral and systemic diseases. In contrast, virus-specific antibodies, which appear during and after hepatitis virus infection, are serologic markers of diagnostic importance.

As described above, serologic tests are available routinely with which to establish a diagnosis of hepatitis A, B, D, and C. Tests for fecal or serum HAV are not routinely available. Therefore, a diagnosis of hepatitis A is based on detection of IgM anti-HAV during acute illness (Fig. 339-2). Rheumatoid factor can give rise to false-positive results in this test.

A diagnosis of HBV infection can usually be made by detection of HBsAg in serum. Infrequently, levels of HBsAg are too low to be detected during acute HBV infection, even with contemporary, highly sensitive immunoassays. In such cases, the diagnosis can be established by the presence of IgM anti-HBc.

The titer of HBsAg bears little relation to the severity of clinical disease. Indeed, an inverse correlation exists between the serum concentration of HBsAg and the degree of liver cell damage. For example, titers are highest in immunosuppressed patients, lower in patients with chronic liver disease (but higher in mild chronic than in severe chronic hepatitis), and very low in patients with acute fulminant hepatitis.

These observations suggest that in hepatitis B the degree of liver cell damage and the clinical course are related to variations in the patient's immune response to HBV rather than to the amount of circulating HBsAg. In immunocompetent persons, however, a correlation exists between markers of HBV replication and liver injury (see below).

Another important serologic marker in patients with hepatitis B is HBeAg. Its principal clinical usefulness is as an indicator of relative infectivity. Because HBeAg is invariably present during early acute hepatitis B, HBeAg testing is indicated primarily in chronic infection.

In patients with hepatitis B surface antigenemia of unknown duration (e.g., blood donors found to be HBsAg-positive) testing for IgM anti-HBc may be useful to distinguish between acute or recent infection (IgM anti-HBc-positive) and chronic HBV infection (IgM anti-HBc-negative, IgG anti-HBc-positive). A false-positive test for IgM anti-HBc may be encountered in patients with high-titer rheumatoid factor. Also, IgM anti-HBc may be reexpressed during acute reactivation of chronic hepatitis B.

Anti-HBs is rarely detectable in the presence of HBsAg in patients with acute hepatitis B, but 10–20% of persons with chronic HBV infection may harbor low-level anti-HBs. This antibody is directed not against the common group determinant, *a*, but against the heterotypic subtype determinant (e.g., HBsAg of subtype *ad* with anti-HBs of subtype *y*). In most cases, this serologic pattern cannot be attributed to infection with two different HBV subtypes but, instead, is thought (based on the clonal selection theory of antibody diversity) to reflect the stimulation of a related clone of antibody-forming cells and is not a harbinger of imminent HBsAg clearance. When such antibody is detected, its presence is of no recognized clinical significance (see "Virology and Etiology").

After immunization with hepatitis B vaccine, which consists of HBsAg alone, anti-HBs is the only serologic marker to appear. The commonly encountered serologic patterns of hepatitis B and their interpretations are summarized in Table 339-5. Tests for the detection of HBV DNA in liver and serum are now available. Like HBeAg, serum HBV DNA is an indicator of HBV replication, but tests for HBV DNA are more sensitive and quantitative. First-generation hybridization assays for HBV DNA had a sensitivity of 10^5 – 10^6 virions/mL, a relative threshold below which infectivity and liver injury are limited and HBeAg is usually undetectable. Currently, testing for HBV DNA has shifted from insensitive hybridization assays to amplification assays (e.g., the PCR-based assay, which can detect as few as 10 or 100 virions/mL); among the commercially available PCR assays, the most useful are those with the highest sensitivity (5–10 IU/mL) and the largest dynamic range (10^3 – 10^9 IU/mL). With increased

sensitivity, amplification assays remain reactive well below the current 10^3 – 10^4 IU/mL threshold for infectivity and liver injury. These markers are useful in following the course of HBV replication in patients with chronic hepatitis B receiving antiviral chemotherapy (Chap. 341). Except for the early decades of life after perinatally acquired HBV infection (see above), in immunocompetent adults with chronic hepatitis B, a general correlation exists between the level of HBV replication, as reflected by the level of serum HBV DNA, and the degree of liver injury. High-serum HBV DNA levels, increased expression of viral antigens, and necroinflammatory activity in the liver go hand in hand unless immunosuppression interferes with cytolytic T-cell responses to virus-infected cells; reduction of HBV replication with antiviral drugs tends to be accompanied by an improvement in liver histology. Among patients with chronic hepatitis B, high levels of HBV DNA increase the risk of cirrhosis, hepatic decompensation, and hepatocellular carcinoma (see "Complications and Sequelae").

In patients with hepatitis C, an episodic pattern of aminotransferase elevation is common. A specific serologic diagnosis of hepatitis C can be made by demonstrating the presence in serum of anti-HCV. When contemporary immunoassays are used, anti-HCV can be detected in acute hepatitis C during the initial phase of elevated aminotransferase activity and remains detectable after recovery (which is rare) and during chronic infection (common). Nonspecificity can confound immunoassays for anti-HCV, especially in persons with a low prior probability of infection, such as volunteer blood donors, or in persons with circulating rheumatoid factor, which can bind nonspecifically to assay reagents; testing for HCV RNA can be used in such settings to distinguish between true-positive and false-positive anti-HCV determinations. Assays for HCV RNA are the most sensitive tests for HCV infection and represent the "gold standard" in establishing a diagnosis of hepatitis C. HCV RNA can be detected even before acute elevation of aminotransferase activity and before the appearance of anti-HCV in patients with acute hepatitis C. In addition, HCV RNA remains detectable indefinitely, continuously in most but intermittently in some, in patients with chronic hepatitis C (detectable as well in some persons with normal liver tests, i.e., inactive carriers). In the very small minority of patients with hepatitis C who lack anti-HCV, a diagnosis can be supported by detection of HCV RNA. If all these tests are negative and the patient has a well-characterized case of hepatitis after percutaneous exposure to blood or blood products, a diagnosis of hepatitis caused by an unidentified agent can be entertained.

Amplification techniques are required to detect HCV RNA. Currently, such target amplification (i.e., synthesis of multiple copies of the viral genome) is achieved by PCR, in which the viral RNA is reverse transcribed to complementary DNA and then amplified by repeated cycles of DNA synthesis. Quantitative PCR assays provide a measurement of relative "viral load"; current PCR assays have a sensitivity of 10 (lower limit of detection) to 25 (lower limit of quantitation) IU/mL and a wide dynamic range (10^3 – 10^9 IU/mL). Determination of HCV RNA level is not a reliable marker of disease severity or prognosis but is helpful in predicting relative responsiveness to antiviral therapy. The same is true for determinations of HCV genotype (Chap. 341). Of course, HCV RNA monitoring during and after antiviral therapy is the *sine qua non* for determining on-treatment and durable responsiveness.

A proportion of patients with hepatitis C have isolated anti-HBc in their blood, a reflection of a common risk in certain populations of exposure to multiple bloodborne hepatitis agents. The anti-HBc in such cases is almost invariably of the IgG class and usually represents HBV infection in the remote past (HBV DNA undetectable); it rarely represents

TABLE 339-5 Commonly Encountered Serologic Patterns of Hepatitis B Infection

HBsAg	ANTI-HBs	ANTI-HBc	HBeAg	ANTI-HBc	INTERPRETATION
+	–	IgM	+	–	Acute hepatitis B, high infectivity ^a
+	–	IgG	+	–	Chronic hepatitis B, high infectivity
+	–	IgG	–	+	1. Late acute or chronic hepatitis B, low infectivity 2. HBeAg-negative ("precore-mutant") hepatitis B (chronic or, rarely, acute)
+	+	+	+/-	+/-	1. HBsAg of one subtype and heterotypic anti-HBs (common) 2. Process of seroconversion from HBsAg to anti-HBs (rare)
–	–	IgM	+/-	+/-	1. Acute hepatitis B ^a 2. Anti-HBc "window"
–	–	IgG	–	+/-	1. Low-level hepatitis B carrier 2. Hepatitis B in remote past
–	+	IgG	–	+/-	Recovery from hepatitis B
–	+	–	–	–	1. Immunization with HBsAg (after vaccination) 2. Hepatitis B in the remote past (?) 3. False-positive

^aIgM anti-HBc may reappear during acute reactivation of chronic hepatitis B.

Note: See text for abbreviations.

current HBV infection with low-level virus carriage. Detectable anti-HCV in the absence of HCV RNA signifies spontaneous or therapeutically induced recovery from (“cured”) hepatitis C.

The presence of HDV infection can be identified by demonstrating intrahepatic HDV antigen or, more practically, an anti-HDV seroconversion (a rise in titer of anti-HDV or de novo appearance of anti-HDV). Circulating HDV antigen, also diagnostic of acute infection, is detectable only briefly, if at all. Because anti-HDV is often undetectable once HBsAg disappears, retrospective serodiagnosis of acute self-limited, simultaneous HBV and HDV infection is difficult. Early diagnosis of acute infection may be hampered by a delay of up to 30–40 days in the appearance of anti-HDV.

When a patient presents with acute hepatitis and has HBsAg and anti-HDV in serum, determination of the class of anti-HBc is helpful in establishing the relationship between infection with HBV and HDV. Although IgM anti-HBc does not distinguish *absolutely* between acute and chronic HBV infection, its presence is a reliable indicator of recent infection and its absence a reliable indicator of infection in the remote past. In simultaneous acute HBV and HDV infections, IgM anti-HBc will be detectable, whereas in acute HDV infection superimposed on chronic HBV infection, anti-HBc will be of the IgG class. Assays for HDV RNA, available in specialized laboratories and yet to be standardized, can be used to confirm HDV infection and to monitor treatment during chronic infection.

The serologic/virologic course of events during acute hepatitis E is entirely analogous to that of acute hepatitis A, with brief fecal shedding of virus and viremia and an early IgM anti-HEV response that predominates during approximately the first 3 months but is eclipsed thereafter by long-lasting IgG anti-HEV. Diagnostic tests of varying reliability for hepatitis E are commercially available outside the United States; in the United States, although tests for HEV infection are not approved by the FDA, reliable diagnostic serologic/virologic assays can be performed at the CDC or other commercial or academic laboratories.

Liver biopsy is rarely necessary or indicated in acute viral hepatitis, except when the diagnosis is questionable or when clinical evidence suggests a diagnosis of chronic hepatitis.

A diagnostic algorithm can be applied in the evaluation of cases of acute viral hepatitis. A patient with acute hepatitis should undergo four serologic tests: HBsAg, IgM anti-HAV, IgM anti-HBc, and anti-HCV (Table 339-6). The presence of HBsAg, with or without IgM anti-HBc, represents HBV infection. If IgM anti-HBc is present, the HBV infection is considered acute; if IgM anti-HBc is absent, the HBV infection is considered chronic. A diagnosis of acute hepatitis B can be made in the absence of HBsAg when IgM anti-HBc is detectable. A diagnosis of acute hepatitis A is based on the presence of IgM anti-HAV. If IgM anti-HAV coexists with HBsAg, a diagnosis of simultaneous HAV and HBV infections can be made; if IgM anti-HBc (with or without HBsAg) is detectable, the patient has simultaneous acute hepatitis A and B, and if IgM anti-HBc is undetectable, the patient has acute hepatitis A superimposed on chronic HBV infection. The presence of anti-HCV supports a diagnosis of acute hepatitis C. Occasionally, testing for HCV RNA or repeat anti-HCV testing later during the illness is necessary to

establish the diagnosis. Absence of all serologic markers is consistent with a diagnosis of “non-A, non-B, non-C” hepatitis (no other proven human hepatitis viruses have been identified), if the epidemiologic setting is appropriate.

In patients with chronic hepatitis, initial testing should consist of HBsAg and anti-HCV. Anti-HCV supports and HCV RNA testing establishes the diagnosis of chronic hepatitis C. If a serologic diagnosis of chronic hepatitis B is made, testing for HBeAg and anti-HBe is indicated to evaluate relative infectivity. Testing for HBV DNA in such patients provides a more quantitative and sensitive measure of the level of virus replication and therefore is very helpful during antiviral therapy (Chap. 341). In patients with chronic hepatitis B and normal aminotransferase activity in the absence of HBeAg, serial testing over time is often required to distinguish between inactive carriage and HBeAg-negative chronic hepatitis B with fluctuating virologic and necroinflammatory activity. In persons with hepatitis B, testing for anti-HDV is useful in those with severe and fulminant disease, with severe chronic disease, with chronic hepatitis B and acute hepatitis-like exacerbations, with frequent percutaneous exposures, and from areas where HDV infection is endemic.

■ PROGNOSIS

Virtually all previously healthy patients with hepatitis A recover completely with no clinical sequelae. Similarly, in acute hepatitis B, 95–99% of previously healthy adults have a favorable course and recover completely. Certain clinical and laboratory features, however, suggest a more complicated and protracted course. Patients of advanced age and with serious underlying medical disorders may have a prolonged course and are more likely to experience severe hepatitis. Initial presenting features such as ascites, peripheral edema, and symptoms of hepatic encephalopathy suggest a poorer prognosis. In addition, a prolonged PT, low serum albumin level, hypoglycemia, and very high serum bilirubin values suggest severe hepatocellular disease. Patients with these clinical and laboratory features deserve prompt hospital admission. The case-fatality rate in hepatitis A and B is very low (~0.1%) but is increased by advanced age and underlying debilitating disorders. Among patients ill enough to be hospitalized for acute hepatitis B, the fatality rate is 1%. Hepatitis C is less severe during the acute phase than hepatitis B and is more likely to be anicteric; fatalities are rare, but the precise case-fatality rate is not known. In outbreaks of waterborne hepatitis E in India and Asia, the case-fatality rate is 1–2% and up to 10–20% in pregnant women. Contributing to fulminant hepatitis E in endemic countries (but only very rarely or not at all in nonendemic countries) are instances of acute hepatitis E superimposed on underlying chronic liver disease (“acute-on-chronic” liver disease). Patients with simultaneous acute hepatitis B and D do not necessarily experience a higher mortality rate than do patients with acute hepatitis B alone; however, in several outbreaks of acute simultaneous HBV and HDV infection among injection drug users, the case-fatality rate was ~5%. When HDV superinfection occurs in a person with chronic hepatitis B, the likelihood of fulminant hepatitis and death is increased substantially. Although the case-fatality rate for hepatitis D is not

known definitively, in outbreaks of severe HDV superinfection in isolated populations with a high hepatitis B carrier rate (“Lábrea fever”), a mortality rate >20% has been recorded.

■ COMPLICATIONS AND SEQUELAE

A small proportion of patients with hepatitis A experience *relapsing hepatitis* weeks to months after apparent recovery from acute hepatitis. Relapses are characterized by recurrence of symptoms, aminotransferase elevations, occasional jaundice, and fecal excretion of HAV. Another unusual variant of acute hepatitis A is *cholestatic hepatitis*, characterized by protracted

TABLE 339-6 Simplified Diagnostic Approach in Patients Presenting with Acute Hepatitis

SEROLOGIC TESTS OF PATIENT'S SERUM				DIAGNOSTIC INTERPRETATION
HBsAg	IgM ANTI-HAV	IgM ANTI-HBc	ANTI-HCV	
+	-	+	-	Acute hepatitis B
+	-	-	-	Chronic hepatitis B
+	+	-	-	Acute hepatitis A superimposed on chronic hepatitis B
+	+	+	-	Acute hepatitis A and B
-	+	-	-	Acute hepatitis A
-	+	+	-	Acute hepatitis A and B (HBsAg below detection threshold)
-	-	+	-	Acute hepatitis B (HBsAg below detection threshold)
-	-	-	+	Acute hepatitis C

Note: See text for abbreviations.

cholestatic jaundice and pruritus. Rarely, liver test abnormalities persist for many months, even up to 1 year. Even when these complications occur, hepatitis A remains self-limited and does not progress to chronic liver disease. During the prodromal phase of acute hepatitis B, a serum sickness-like syndrome characterized by arthralgia or arthritis, rash, angioedema, and, rarely, hematuria and proteinuria may develop in 5–10% of patients. This syndrome occurs before the onset of clinical jaundice, and these patients are often diagnosed erroneously as having rheumatologic diseases. The diagnosis can be established by measuring serum aminotransferase levels, which are almost invariably elevated, and serum HBsAg. As noted above, EMC is an immune-complex disease that can complicate chronic hepatitis C and is part of a spectrum of B-cell lymphoproliferative disorders, which, in rare instances, can evolve to B-cell lymphoma (Chap. 108). Attention has been drawn as well to associations between hepatitis C and such cutaneous disorders as porphyria cutanea tarda and lichen planus. A mechanism for these associations is unknown. Related to the reliance of HCV on lipoprotein secretion and assembly pathways and on interactions of HCV with glucose metabolism, HCV infection may be complicated by hepatic steatosis, hypercholesterolemia, insulin resistance (and other manifestations of the metabolic syndrome), and type 2 diabetes mellitus; both hepatic steatosis and insulin resistance appear to accelerate hepatic fibrosis and blunt responsiveness to interferon-based antiviral therapy (Chap. 341). Finally, chronic hepatitis C has been linked to multiple extrahepatic disorders, including cardiovascular and cerebrovascular disease, renal disease, rheumatologic/immunologic disorders, mental health and cognitive disorders (many patients describe “brain fog”), and, in addition to hepatocellular carcinoma, nonliver malignancies.

The most feared complication of viral hepatitis is *fulminant hepatitis* (massive hepatic necrosis); fortunately, this is a rare event. Fulminant hepatitis is seen primarily in hepatitis B, D, and E, but rare fulminant cases of hepatitis A occur primarily in older adults and in persons with underlying chronic liver disease, including, according to some reports, chronic hepatitis B and C. Hepatitis B accounts for >50% of fulminant cases of viral hepatitis, a sizable proportion of which are associated with HDV infection and another proportion with underlying chronic hepatitis C. Fulminant hepatitis is hardly ever seen in hepatitis C, but hepatitis E, as noted above, can be complicated by fatal fulminant hepatitis in 1–2% of all cases and in up to 20% of cases in pregnant women. Patients usually present with signs and symptoms of encephalopathy that may evolve to deep coma. The liver is usually small and the PT excessively prolonged. The combination of rapidly shrinking liver size, rapidly rising bilirubin level, and marked prolongation of the PT, even as aminotransferase levels fall, together with clinical signs of confusion, disorientation, somnolence, ascites, and edema, indicates that the patient has hepatic failure with encephalopathy. Cerebral edema is common; brainstem compression, gastrointestinal bleeding, sepsis, respiratory failure, cardiovascular collapse, and renal failure are terminal events. The mortality rate is exceedingly high (>80% in patients with deep coma), but patients who survive may have a complete biochemical and histologic recovery. If a donor liver can be located in time, liver transplantation may be lifesaving in patients with fulminant hepatitis (Chap. 345).

Documenting the disappearance of HBsAg after apparent clinical recovery from acute hepatitis B is particularly important. Before laboratory methods were available to distinguish between acute hepatitis and acute hepatitis-like exacerbations (*spontaneous reactivations*) of chronic hepatitis B, observations suggested that ~10% of previously healthy patients remained HBsAg positive for >6 months after the onset of clinically apparent acute hepatitis B. One-half of these persons cleared the antigen from their circulations during the next several years, but the other 5% remained chronically HBsAg positive. More recent observations suggest that the true rate of chronic infection after clinically apparent acute hepatitis B is as low as 1% in normal, immunocompetent, young adults. Earlier, higher estimates may have been confounded by inadvertent inclusion of acute exacerbations in chronically infected patients; these patients, chronically HBsAg positive before exacerbation, were unlikely to seroconvert to HBsAg negative thereafter. Whether the rate of chronicity is 10% or 1%, such patients

have IgG anti-HBc in serum; anti-HBs is either undetected or detected at low titer against the opposite subtype specificity of the antigen (see “*Laboratory Features*”). These patients may (1) be inactive carriers; (2) have low-grade, mild chronic hepatitis; or (3) have moderate to severe chronic hepatitis with or without cirrhosis. The likelihood of remaining chronically infected after acute HBV infection is especially high among neonates, persons with Down’s syndrome, chronically hemodialyzed patients, and immunosuppressed patients, including persons with HIV infection.

Chronic hepatitis is an important late complication of acute hepatitis B occurring in a small proportion of patients with acute disease but more common in those who present with chronic infection without having experienced an acute illness, as occurs typically after neonatal infection or after infection in an immunosuppressed host (Chap. 341). The following clinical and laboratory features suggest progression of acute hepatitis to chronic hepatitis: (1) lack of complete resolution of clinical symptoms of anorexia, weight loss, fatigue, and the persistence of hepatomegaly; (2) the presence of bridging/interface or multilobular hepatic necrosis on liver biopsy during protracted, severe acute viral hepatitis; (3) failure of the serum aminotransferase, bilirubin, and globulin levels to return to normal within 6–12 months after the acute illness; and (4) the persistence of HBeAg for >3 months or HBsAg for >6 months after acute hepatitis.

Although acute hepatitis D infection does not increase the likelihood of chronicity of simultaneous acute hepatitis B, hepatitis D has the potential for contributing to the severity of chronic hepatitis B. Hepatitis D superinfection can transform inactive or mild chronic hepatitis B into severe, progressive chronic hepatitis and cirrhosis; it also can accelerate the course of chronic hepatitis B and accelerate the risk of hepatocellular carcinoma. Some HDV superinfections in patients with chronic hepatitis B lead to fulminant hepatitis. As defined in longitudinal studies over three decades, the annual rate of cirrhosis in patients with chronic hepatitis D is 4%. Although HDV and HBV infections are associated with severe liver disease, mild hepatitis and even inactive carriage have been identified in some patients, and the disease may become indolent beyond the early years of infection.

After acute HCV infection, the likelihood of remaining chronically *infected* approaches 85–90%. Although many patients with chronic hepatitis C have no symptoms, cirrhosis may develop in as many as 20% within 10–20 years of acute illness; in some series of cases reported by referral centers, cirrhosis has been reported in as many as 50% of patients with chronic hepatitis C. Among cirrhotic patients with chronic hepatitis C, the annual risk of hepatic decompensation is ~4%. Although prior to the availability of highly effective DAA therapy during the second decade of the twenty-first century chronic hepatitis C accounted for at least 40% of cases of chronic liver disease and of patients undergoing liver transplantation for end-stage liver disease in the United States and Europe, in the majority of patients with chronic hepatitis C, morbidity and mortality are limited during the initial 20 years after the onset of infection. Progression of chronic hepatitis C may be influenced by advanced age of acquisition, long duration of infection, immunosuppression, coexisting excessive alcohol use, concomitant hepatic steatosis, other hepatitis virus infection, or HIV co-infection. In fact, instances of severe and rapidly progressive chronic hepatitis B and C are being recognized with increasing frequency in patients with HIV infection (Chap. 202). In contrast, neither HAV nor HEV causes chronic liver disease in immunocompetent hosts; however, cases of chronic hepatitis E (including cirrhosis and end-stage liver disease and even hepatocellular carcinoma) have been observed in immunosuppressed organ-transplant recipients, persons receiving cytotoxic chemotherapy, and persons with HIV infection. Among patients with chronic hepatitis (e.g., caused by hepatitis B or C, alcohol, etc.) in endemic countries, hepatitis E has been reported as the cause of acute-on-chronic liver failure; however, in most experiences among patients from nonendemic countries, HEV has not been found to contribute commonly to hepatic decompensation in patients with chronic hepatitis.

Persons with chronic hepatitis B, particularly those infected in infancy or early childhood and especially those with HBeAg and/or

high-level HBV DNA, have an enhanced risk of hepatocellular carcinoma. The risks of cirrhosis and hepatocellular carcinoma increase with the level of HBV replication. The annual rate of hepatocellular carcinoma in patients with chronic hepatitis D and cirrhosis is ~3%. The risk of hepatocellular carcinoma is increased as well in patients with chronic hepatitis C, almost exclusively in patients with cirrhosis, and almost always after at least several decades, usually after three decades of disease (Chap. 82). Among such cirrhotic patients with chronic hepatitis C, the annual risk of hepatocellular carcinoma is ~1–4%.

Rare complications of viral hepatitis include pancreatitis, myocarditis, atypical pneumonia, aplastic anemia, transverse myelitis, and peripheral neuropathy. In children, hepatitis B may present rarely with anicteric hepatitis, a nonpruritic papular rash of the face, buttocks, and limbs, and lymphadenopathy (papular acrodermatitis of childhood or Gianotti-Crosti syndrome).

Rarely, autoimmune hepatitis (Chap. 341) can be triggered by a bout of otherwise self-limited acute hepatitis, as reported after acute hepatitis A, B, and C.

■ DIFFERENTIAL DIAGNOSIS

Viral diseases such as infectious mononucleosis; those due to cytomegalovirus, herpes simplex, and coxsackieviruses; and toxoplasmosis may share certain clinical features with viral hepatitis and cause elevations in serum aminotransferase and, less commonly, in serum bilirubin levels. Tests such as the differential heterophile and serologic tests for these agents may be helpful in the differential diagnosis if HBsAg, anti-HBc, IgM anti-HAV, and anti-HCV determinations are negative. Aminotransferase elevations can accompany almost any systemic viral infection, including the coronavirus SARS-CoV-2 (~10% of all cases and up to half of severe cases); other rare causes of liver injury confused with viral hepatitis are infections with *Leptospira*, *Candida*, *Brucella*, *Mycobacteria*, and *Pneumocystis*. A complete drug history is particularly important because many drugs and certain anesthetic agents can produce a picture of either acute hepatitis or cholestasis (Chap. 340). Equally important is a history of unexplained “repeated episodes” of acute hepatitis. This history should alert the physician to the possibility that the underlying disorder is chronic hepatitis, for example, autoimmune hepatitis (Chap. 341). Alcoholic hepatitis must also be considered, but usually the serum aminotransferase levels are not as markedly elevated, and other stigmata of alcoholism may be present. The finding on liver biopsy of fatty infiltration, a neutrophilic inflammatory reaction, and “alcoholic hyaline” would be consistent with alcohol-induced rather than viral liver injury. Because acute hepatitis may present with right upper quadrant abdominal pain, nausea and vomiting, fever, and icterus, it is often confused with acute cholecystitis, common duct stone, or ascending cholangitis. Patients with acute viral hepatitis may tolerate surgery poorly; therefore, it is important to exclude this diagnosis, and in confusing cases, a percutaneous liver biopsy may be necessary before laparotomy. Viral hepatitis in the elderly is often misdiagnosed as obstructive jaundice resulting from a common duct stone or carcinoma of the pancreas. Because acute hepatitis in the elderly may be quite severe and the operative mortality high, a thorough evaluation including biochemical tests, radiographic studies of the biliary tree, and even liver biopsy may be necessary to exclude primary parenchymal liver disease. Another clinical constellation that may mimic acute hepatitis is right ventricular failure with passive hepatic congestion or hypoperfusion syndromes, such as those associated with shock, severe hypotension, and severe left ventricular failure. Also included in this general category is any disorder that interferes with venous return to the heart, such as right atrial myxoma, constrictive pericarditis, hepatic vein occlusion (Budd-Chiari syndrome), or veno-occlusive disease. Clinical features are usually sufficient to distinguish among these vascular disorders and viral hepatitis. Acute fatty liver of pregnancy, cholestasis of pregnancy, eclampsia, and the HELLP (*hemolysis*, *elevated liver tests*, and *low platelets*) syndrome can be confused with viral hepatitis during pregnancy. Very rarely, malignancies metastatic to the liver can mimic acute or even fulminant viral hepatitis. Occasionally, genetic or metabolic liver disorders (e.g.,

Wilson’s disease, α_1 antitrypsin deficiency) and nonalcoholic fatty liver disease are confused with acute viral hepatitis. Among patients with biochemical evidence for severe liver injury, i.e., aminotransferase levels of ≥ 1000 IU/L, the most common causes are ischemic liver injury, drug-induced liver injury (especially caused by acetaminophen), acute viral hepatitis, and pancreaticobiliary disorders.

TREATMENT

Acute Viral Hepatitis

Most persons with acute hepatitis (especially hepatitis A, B, and E) recover spontaneously and do not require specific antiviral therapy. In hepatitis B, among previously healthy adults who present with clinically apparent acute hepatitis, recovery occurs in ~99%; therefore, antiviral therapy is not likely to improve the rate of recovery and is not required. In rare instances of severe acute hepatitis B, treatment with a nucleoside analogue at oral doses used to treat chronic hepatitis B (Chap. 341) has been attempted successfully. Although clinical trials have not been done to establish the efficacy or duration of this approach, most authorities would recommend institution of antiviral therapy with a nucleoside analogue (entecavir or tenofovir, the most potent and least resistance-prone agents) for severe, but not mild-moderate, acute hepatitis B. Treatment should continue until 3 months after HBsAg seroconversion or 6 months after HBeAg seroconversion.

In typical cases of acute hepatitis C, recovery is rare (~15–20% in most experiences), and progression to chronic hepatitis is the rule. Patients with jaundice, those with HCV genotype 1, women, and those with earlier age of infection, lower level of HCV RNA, HBV co-infection, and absence of current injection drug use are more likely to recover from acute hepatitis C, as are persons who have genetic markers associated with spontaneous recovery (*IL28B* CC haplotype).

Because spontaneous recovery can occur and because most cases of acute hepatitis C are not clinically severe or rapidly progressive, delaying antiviral therapy of acute hepatitis C for 3–6 months (after which recovery is unlikely) was a recommended approach during the era of interferon-based therapy; however, in the current era of highly effective (95–100%) oral DAA therapy, waiting for potential spontaneous recovery is no longer advised; instead, early treatment with one of the four first-line drug combinations (of polymerase inhibitors, protease inhibitors, and/or NSSA inhibitors) approved for treatment of chronic hepatitis C (Chap. 341) is recommended for treatment of patients with acute hepatitis C. Although abbreviated treatment courses have been studied, currently, a standard, full 8- to 12-week course is recommended.

Because of the vast reservoir of acute HCV infections acquired four to five decades ago in the 1945–1965 birth cohort, most newly recognized HCV infections are chronic. Opportunities to identify and treat patients with acute hepatitis C occur in two population subsets: (1) in health care workers who sustain hepatitis C–contaminated needle sticks (occupational accidents), monitoring for ALT elevations and the presence of HCV RNA identify acute hepatitis C in ~3%, and this group should be treated; (2) in injection drug users, the risk of acute hepatitis C has been on the rise during the previous decade, and the epidemic of opioid use has contributed to an amplification of HCV infection among drug users. Such patients are candidates for antiviral therapy, and efforts to combine antiviral therapy with drug rehabilitation therapy have been very successful.

Notwithstanding these specific therapeutic considerations, in most cases of typical acute viral hepatitis, specific treatment generally is not necessary. Although hospitalization may be required for clinically severe illness, most patients do not require hospital care. Forced and prolonged bed rest is not essential for full recovery, but many patients will feel better with restricted physical activity. A high-calorie diet is desirable, and because many patients may experience nausea late in the day, the major caloric intake is best tolerated in the morning. Intravenous feeding is necessary in the

acute stage if the patient has persistent vomiting and cannot maintain oral intake. Drugs capable of producing adverse reactions such as cholestasis and drugs metabolized by the liver should be avoided. If severe pruritus is present, the use of the bile salt-sequestering resin cholestyramine is helpful. Glucocorticoid therapy has no value in acute viral hepatitis, even in severe cases, and may be deleterious, even increasing the risk of chronicity (e.g., of acute hepatitis B).

Physical isolation of patients with hepatitis to a single room and bathroom is rarely necessary except in the case of fecal incontinence for hepatitis A and E or uncontrolled, voluminous bleeding for hepatitis B (with or without concomitant hepatitis D) and C. Because most patients hospitalized with hepatitis A excrete little, if any, HAV, the likelihood of HAV transmission from these patients during their hospitalization is low. Therefore, burdensome *enteric precautions are no longer recommended*. Although gloves should be worn when the bed pans or fecal material of patients with hepatitis A are handled, these precautions do not represent a departure from sensible procedure and contemporary universal precautions for all hospitalized patients. For patients with hepatitis B and C, emphasis should be placed on blood precautions (i.e., avoiding direct, ungloved hand contact with blood and other body fluids). Enteric precautions are unnecessary. The importance of simple hygienic precautions such as hand washing cannot be overemphasized. Universal precautions that have been adopted for all patients apply to patients with viral hepatitis. Hospitalized patients may be discharged following substantial symptomatic improvement, a significant downward trend in the serum aminotransferase and bilirubin values, and a return to normal of the PT. Mild aminotransferase elevations should not be considered contraindications to the gradual resumption of normal activity.

In *fulminant hepatitis*, the goal of therapy is to support the patient by maintenance of fluid balance, support of circulation and respiration, control of bleeding, correction of hypoglycemia, and treatment of other complications of the comatose state in anticipation of liver regeneration and repair. Protein intake should be restricted, and oral lactulose administered. Glucocorticoid therapy has been shown in controlled trials to be ineffective. Likewise, exchange transfusion, plasmapheresis, human cross-circulation, porcine liver cross-perfusion, hemoperfusion, and extracorporeal liver-assist devices have not been proven to enhance survival. Meticulous intensive care that includes prophylactic antibiotic coverage is the one factor that appears to improve survival. Orthotopic liver transplantation is resorted to with increasing frequency, with excellent results, in patients with fulminant hepatitis (Chap. 345). Fulminant hepatitis C is very rare; however, in fulminant hepatitis B, oral antiviral therapy has been used successfully, as reported anecdotally. In clinically severe acute hepatitis E or acute-on-chronic liver failure, successful therapy with ribavirin (600 mg twice daily, 15 mg/kg) has been reported anecdotally. Unfortunately, when fulminant hepatitis E occurs in pregnant women (as it does in up to 20% of pregnant women with acute hepatitis E), ribavirin, which is teratogenic, is contraindicated. In cases of hepatitis E in organ-transplant recipients, reduction in overall immunosuppressive drug doses and switching from tacrolimus to cyclosporine A have been shown to be effective, often without antiviral therapy, in achieving eradication of HEV. If a change in immunosuppression is inadequate, ribavirin treatment for 3 months has been observed to achieve a sustained virologic response in 78% of treated patients; however, the optimal dose and duration of ribavirin therapy remain to be determined.

■ PROPHYLAXIS

Because application of therapy for acute viral hepatitis is limited and because chronic viral hepatitis requires prolonged and costly courses of antiviral therapy (Chap. 341), emphasis is placed on prevention through immunization. The prophylactic approach differs for each of the types of viral hepatitis. In the past, immunoprophylaxis relied exclusively on passive immunization with antibody-containing

globulin preparations purified by cold ethanol fractionation from the plasma of hundreds of normal donors. Currently, for hepatitis A, B, and E, active immunization with vaccines is the preferable approach to prevention.

Hepatitis A Both passive immunization with immunoglobulin (IG) and active immunization with killed vaccines are available. All preparations of IG contain anti-HAV concentrations sufficient to be protective. Administration of plasma-derived globulin is safe; all contemporary lots of IG are subjected to viral inactivation steps and must be free of HCV RNA as determined by PCR testing. Administration of IM lots of IG has not been associated with transmission of HBV, HCV, or HIV. When administered before exposure or during the early incubation period, IG is effective in preventing clinically apparent hepatitis A. For postexposure prophylaxis of intimate contacts (household, sexual, institutional) of persons with hepatitis A, the administration of 0.02 mL/kg is recommended as early after exposure as possible; it may be effective even when administered as late as 2 weeks after exposure. Prophylaxis is not necessary for those who have already received hepatitis A vaccine, for casual contacts (office, factory, school, or hospital), for most elderly persons, who are very likely to be immune, or for those known to have anti-HAV in their serum. By the time most common-source outbreaks of hepatitis A are recognized, it is usually too late in the incubation period for IG to be effective; however, prophylaxis may have limited the frequency of secondary cases. For travelers to tropical countries, developing countries, and other areas outside standard tourist routes, IG prophylaxis had been recommended before a vaccine became available.

Such IG recommendations for postexposure prophylaxis and for preexposure prophylaxis for international travel were updated in 2018. Currently, hepatitis A vaccine, not IG, is recommended for all persons aged ≥ 12 months for postexposure prophylaxis and for preexposure prophylaxis prior to international travel to HAV-endemic areas. For adults aged >40 , IG (at an upward revised dose of 0.1 mg/kg) may be added to postexposure hepatitis B vaccination depending on an assessment of the person's risk. Even though hepatitis A vaccine is indicated for children ≥ 12 months of age, when infants aged 6–11 months travel internationally to areas with a risk of HAV infection, they should receive the vaccine for preexposure prophylaxis; however, this travel-related dose should not be counted toward the universal childhood two-dose hepatitis A vaccine recommendation, which begins at age 12 months. For postexposure prophylaxis of persons with contraindications to hepatitis A vaccination and infants aged <12 months, the use of IG (0.1 mL/kg) should be retained. In addition, for postexposure prophylaxis in immunocompromised adults and persons with chronic liver disease, both hepatitis A vaccination and IG administration (0.1 mL/kg), at different IM sites, are recommended. Finally, for infants aged <6 months and for persons with contraindications to hepatitis A vaccination, preexposure prophylaxis for travel consists of IG at doses of 0.1 mg/kg for travel durations up to 1 month, 0.2 mg/kg for travel up to 2 months, and repeat 0.2 mg/kg every 2 months thereafter for the remainder of travel. Thus, except for these limited considerations, hepatitis A vaccine has supplanted IG in almost all cases for both postexposure prophylaxis and preexposure prophylaxis for travel. Unlike IG prophylaxis, the protection afforded by active immunization with vaccine is durable and simpler to administer.

Formalin-inactivated vaccines made from strains of HAV attenuated in tissue culture have been shown to be safe, immunogenic, and effective in preventing hepatitis A. Hepatitis A vaccines are approved for use in persons who are at least 1 year old and appear to provide adequate protection beginning 4 weeks after a primary inoculation. As noted above, for travel to an endemic area, hepatitis A vaccine is the preferred approach to *preexposure* immunoprophylaxis and provides long-lasting protection (protective levels of anti-HAV should last at least 20 years after vaccination). Shortly after its introduction, hepatitis A vaccine was recommended for children living in communities with a high incidence of HAV infection; in 1999, this recommendation was extended to include all children living in states, counties, and

TABLE 339-7 Hepatitis A Vaccination Schedules

AGE, YEARS	NO. OF DOSES	DOSE	SCHEDULE, MONTHS
HAVERIX (GlaxoSmithKline)^a			
1–18	2	720 ELU ^b (0.5 mL)	0, 6–12
≥19	2	1440 ELU (1 mL)	0, 6–12
VAQTA (Merck)			
1–18	2	25 units (0.5 mL)	0, 6–18
≥19	2	50 units (1 mL)	0, 6–18

^aA combination of this hepatitis A vaccine and hepatitis B vaccine, TWINRIX, is licensed for simultaneous protection against both of these viruses among adults (age ≥18 years). Each 1-mL dose contains 720 ELU of hepatitis A vaccine and 20 µg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6.

^bEnzyme-linked immunoassay units. ^cCombination hepatitis A and typhoid vaccines, Hepatyrix (GlaxoSmithKline) and Viatim (Sanofi Pasteur), are available, targeted primarily for travelers to endemic areas. Please consult product insert for doses and schedules.

communities with high rates of HAV infection. As of 2006, the Advisory Committee on Immunization Practices of the U.S. Public Health Service recommended *routine hepatitis A vaccination of all children*. Other groups considered being at increased risk for HAV infection and who are candidates for hepatitis A vaccination include military personnel, populations with cyclic outbreaks of hepatitis A (e.g., Alaskan natives), employees of day-care centers and persons working in facilities for the developmentally delayed, primate handlers, laboratory workers exposed to hepatitis A or fecal specimens, and patients with chronic liver disease (including persons with aminotransferase elevations ≥2 times the upper limit of normal). Because of an increased risk of fulminant hepatitis A—observed in some experiences but not confirmed in others—among patients with chronic hepatitis C, patients with chronic hepatitis C are candidates for hepatitis A vaccination, as are persons with chronic hepatitis B and the expanding population of persons with nonalcoholic liver disease. Other populations whose recognized risk of hepatitis A is increased should be vaccinated, including men who have sex with men, injection or noninjection drug users, persons experiencing homelessness, persons with clotting disorders who require frequent administration of clotting-factor concentrates, persons traveling from the United States to countries with high or intermediate hepatitis A endemicity, postexposure prophylaxis for contacts of persons with hepatitis A, and household members and other close contacts of adopted children arriving from countries with high and moderate hepatitis A endemicity. Hepatitis A vaccine is now recommended as well for pregnant women at risk of infection or severe outcomes from infection during pregnancy. Recommendations for dose and frequency differ for the two approved vaccine preparations in the United States and the combination vaccines that include hepatitis A (Table 339-7); all injections are IM. Hepatitis A vaccine has been reported to be effective in preventing secondary household and day-care center-associated cases of acute hepatitis A. In the United States, reported mortality resulting from hepatitis A declined in parallel with hepatitis A vaccine-associated reductions in the annual incidence of new infections.

Hepatitis B Until 1982, prevention of hepatitis B was based on *passive immunoprophylaxis* either with standard Ig, containing modest levels of anti-HBs, or hepatitis B immunoglobulin (HBIG), containing high-titer anti-HBs. The efficacy of standard Ig has never been established and remains questionable; even the efficacy of HBIG, demonstrated in several clinical trials, has been challenged, and its contribution appears to be in reducing the frequency of clinical *illness*, not in preventing *infection*. The first vaccine for *active immunization*, introduced in 1982, was prepared from purified, noninfectious, 22-nm spherical HBsAg particles derived from the plasma of healthy HBsAg carriers. In 1987, the plasma-derived vaccine was supplanted by a genetically engineered vaccine derived from recombinant yeast. The latter vaccine consists of HBsAg particles that are nonglycosylated but are otherwise indistinguishable from natural HBsAg; two recombinant vaccines were licensed for use in the United States in the 1980s

(Recombivax-HB 1986; Engerix-B 1989), and a third (Heplisav-B) was licensed in 2017. Current recommendations can be divided into those for preexposure and postexposure prophylaxis.

For *preexposure prophylaxis* against hepatitis B in settings of frequent exposure (health workers exposed to blood; first-responder public safety workers; hemodialysis patients and staff; residents and staff of custodial institutions for the developmentally handicapped; injection drug users; incarcerated inmates of correctional facilities; persons with multiple sexual partners or who have had a sexually transmitted disease; men who have sex with men; persons such as hemophiliacs who require long-term, high-volume therapy with blood derivatives; household and sexual contacts of persons with chronic HBV infection; persons living in or traveling extensively in endemic areas; unvaccinated children aged <18; unvaccinated children who are Alaskan natives, Pacific Islanders, or residents in households of first-generation immigrants from endemic countries; persons born in countries with a prevalence of HBV infection ≥2%; patients with chronic liver disease [including persons with aminotransferase levels >2 times the upper limit of normal]; persons aged <60 years with diabetes mellitus [those ≥60 years at the discretion of their physicians]; persons with end-stage renal disease; and persons with HIV infection), three IM (deltoid, not gluteal) injections of hepatitis B vaccine are recommended at 0, 1, and 6 months (other, optional schedules are summarized in Table 339-8). Pregnancy is *not* a contraindication to vaccination (but Heplisav-B is not recommended for pregnant women because of the lack of safety data in this subpopulation; details of the use of Heplisav-B, a two-injection course a month apart, appear in Table 339-8). In areas of low HBV endemicity such as the United States, despite the availability of safe and effective hepatitis B vaccines, a strategy of vaccinating persons in high-risk groups was not effective. The incidence of new hepatitis B cases continued to increase in the United States after the introduction of vaccines; <10% of all targeted persons in high-risk groups were actually vaccinated, and ~30% of persons with sporadic acute hepatitis B did not fall into any high-risk group category. Therefore, to have an impact on the frequency of HBV infection in an area of low endemicity such as the United States, universal hepatitis B vaccination in childhood is recommended. For unvaccinated children born after the implementation of universal infant vaccination, vaccination during early adolescence, at age 11–12 years, is recommended, and this recommendation has been extended to include all unvaccinated children aged 0–19 years. In HBV-hyperendemic areas (e.g., Asia), universal vaccination of children has resulted in a marked (~70–90%) 30-year decline in complications of hepatitis B, including liver-related mortality and hepatocellular carcinoma.

The original two available aluminum-adjuvanted recombinant hepatitis B vaccines are comparable, one containing 10 µg of HBsAg (Recombivax-HB) and the other containing 20 µg of HBsAg (Engerix-B), and recommended doses for each injection vary for the two preparations (Table 339-8). Combinations of hepatitis B vaccine with other childhood vaccines are available as well (Table 339-8).

In 2017, a third recombinant hepatitis B vaccine with a novel adjuvant that activates Toll-like 9 receptors was approved for adults aged 18 or older. In a series of prospective trials, compared to three Engerix-B injections, two IM doses a month apart yielded higher proportions with protective levels of anti-HBs (≥10 mIU/mL): 95% of adults aged 18–55 or 18–70 (vs 81% for Engerix-B), 90% of older adults aged 40–70 (vs 71% for Engerix-B), and 90% of adults aged 18–70 with type 2 diabetes (vs 65% for Engerix-B). This two-injection regimen may be useful for revaccination of persons who failed to respond to the original vaccines. Another novel recombinant vaccine (PreHevbrio, VBI Vaccines) containing of all three hepatitis B surface antigens, S, pre-S1, and pre-S2, has been shown in clinical trials (three IM doses at 0, 1, and 6 months) to achieve higher proportions with protective anti-HBs and higher antibody levels than Engerix-B (which contains S antigen only), including in older persons (≥45 years), persons with diabetes, and overweight persons (body mass index >30); approved originally outside the United States, this vaccine was approved by the FDA on December 1, 2021 for adults age ≥18 years. Availability is expected during the first quarter of 2022.

TABLE 339-8 Preexposure Hepatitis B Vaccination Schedules

TARGET GROUP	NO. OF DOSES	DOSE	SCHEDULE, MONTHS
Recombivax-HB (Merck)^a			
Infants, children (<1–10 years)	3	5 µg (0.5 mL)	0, 1–2, 4–6
Adolescents (11–19 years)	3 or 4 or	5 µg (0.5 mL)	0–2, 1–4, 4–6 or 0, 12, 24 or 0, 1, 2, 12
Adults (≥20 years)	2	10 µg (1 mL)	0, 4–6 (age 11–15)
Hemodialysis patients ^b	3	10 µg (1 mL)	0–2, 1–4, 4–6
<20 years	3	5 µg (0.5 mL)	0, 1, 6
≥20 years	3	40 µg (4 mL)	0, 1, 6
Engerix-B (GlaxoSmithKline)^c			
Infants, children (<1–10 years)	3 or 4	10 µg (0.5 mL)	0, 1–2, 4–6 or 0, 1, 2, 12
Adolescents (10–19 years)	3 or 4	10 µg (0.5 mL)	0, 1–2, 4–6 or 0, 12, 24 or 0, 1, 2, 12
Adults (≥20 years)	3 or 4	20 µg (1 mL)	0–2, 1–4, 4–6 or 0, 1, 2, 12
Hemodialysis patients ^b			
<20 years	4	10 µg (0.5 mL)	0, 1, 2, 6
≥20 years	4	40 µg (2 mL)	0, 1, 2, 6
Heplisav-B (Dynavax)^d			
Adults (≥18 years)	2	20 µg (0.5 mL)	0, 1

^aThis manufacturer produces a licensed combination of hepatitis B vaccine and vaccines against *Haemophilus influenzae* type b and *Neisseria meningitidis*, Comvax, for use in infants and young children. Please consult product insert for dose and schedule. ^bThis group also includes other immunocompromised persons. ^cThis manufacturer produces two licensed combination hepatitis B vaccines: (1) Twinrix, recombinant hepatitis B vaccine plus inactivated hepatitis A vaccine, is licensed for simultaneous protection against both of these viruses among adults (age ≥18 years). Each 1-mL dose contains 720 ELU (enzyme-linked immunoassay units) of hepatitis A vaccine and 20 µg of hepatitis B vaccine. These doses are recommended at months 0, 1, and 6. (2) Pediarix, recombinant hepatitis B vaccine plus diphtheria and tetanus toxoid, pertussis, and inactivated poliovirus, is licensed for use in infants and young children. A hexavalent vaccine combining diphtheria, tetanus toxoid, pertussis, poliovirus, *H. influenzae* type b, and hepatitis B (Vaxelis, MCM Vaccine Company) was approved by the U.S. Food and Drug Administration in 2018. Please consult product insert for doses and schedules. ^dHeplisav-B has not been tested for safety and efficacy in children, adolescents, hemodialysis patients, and pregnant women; it is not approved for these subpopulations.

For unvaccinated persons sustaining an exposure to HBV, *postexposure* prophylaxis with a combination of HBIG (for rapid achievement of high-titer circulating anti-HBs) and hepatitis B vaccine (for achievement of long-lasting immunity as well as its apparent efficacy in attenuating clinical illness after exposure) is recommended. For *perinatal* exposure of infants born to HBsAg-positive mothers, a single dose of HBIG, 0.5 mL, should be administered IM in the thigh *immediately after birth*, followed by a complete course of three injections of recombinant hepatitis B vaccines approved for children (see doses above) to be started within the first 12 h of life. For those experiencing a direct percutaneous inoculation or transmucosal exposure to HBsAg-positive blood or body fluids (e.g., accidental needle stick, other mucosal penetration, or ingestion), a single IM dose of HBIG, 0.06 mL/kg, administered as soon after exposure as possible, is followed by a complete course of hepatitis B vaccine to begin within the first week. For pregnant mothers with high-level HBV DNA ($>2 \times 10^5$ IU/mL), adding antiviral nucleoside analogues (e.g., pregnancy class B tenofovir, see Chap 341) during the third trimester of pregnancy reduces perinatal transmission even further. For persons exposed by sexual contact to a patient with acute hepatitis B, a single IM dose of HBIG, 0.06 mL/kg, should be given within 14 days of exposure, to be followed by a complete course of hepatitis B vaccine. When both HBIG and hepatitis B vaccine are recommended, they may be given at the same time but at separate sites. Testing adults for anti-HBs after a course of vaccine is advisable to document the acquisition of immunity, but because hepatitis B vaccine immunogenicity is nearly universal in infants, postvaccination anti-HBs testing of children is not recommended.

The precise duration of protection afforded by hepatitis B vaccine is unknown; however, ~80–90% of immunocompetent adult vaccinees retain protective levels of anti-HBs for at least 5 years, and 60–80% for 10 years, and protective antibody has been documented to last for at least two decades after vaccination in infancy. Thereafter and even after anti-HBs becomes undetectable, protection persists

against clinical hepatitis B, hepatitis B surface antigenemia, and chronic HBV infection. Currently, *booster* immunizations are not recommended routinely, except in immunosuppressed persons who have lost detectable anti-HBs or immunocompetent persons who sustain percutaneous HBsAg-positive inoculations after losing detectable antibody. Specifically, for hemodialysis patients, annual anti-HBs testing is recommended after vaccination; booster doses are recommended when anti-HBs levels fall to <10 mIU/mL. As noted above, for persons at risk of both hepatitis A and B, a combined vaccine is available containing 720 enzyme-linked immunoassay units (ELUs) of inactivated HAV and 20 µg of recombinant HBsAg (at 0, 1, and 6 months).

Hepatitis D Infection with hepatitis D can be prevented by vaccinating susceptible persons with hepatitis B vaccine. No product is available for immunoprophylaxis to prevent HDV superinfection in persons with chronic HBV infection; for these patients, avoidance of percutaneous exposures and limitation of intimate contact with persons who have HDV infection are recommended.

Hepatitis C IG is ineffective in preventing hepatitis C and is no longer recommended for postexposure prophylaxis in cases of perinatal, needle stick, or sexual exposure. Although prototype vaccines that induce antibodies to HCV envelope proteins have been developed, currently, hepatitis C vaccination is not feasible practically. Geno-

type and quasispecies viral heterogeneity, as well as rapid evasion of neutralizing antibodies by this rapidly mutating virus, conspire to render HCV a difficult target for immunoprophylaxis with a vaccine. Prevention of transfusion-associated hepatitis C has been accomplished by the following successively introduced measures: exclusion of commercial blood donors and reliance on a volunteer blood supply; screening donor blood with surrogate markers such as ALT (no longer recommended) and anti-HBc, markers that identify segments of the blood donor population with an increased risk of bloodborne infections; exclusion of blood donors in high-risk groups for AIDS and the introduction of anti-HIV screening tests; and progressively sensitive serologic and virologic screening tests for HCV infection.

In the absence of active or passive immunization, prevention of hepatitis C includes behavior changes and precautions to limit exposures to infected persons. Recommendations designed to identify patients with clinically inapparent hepatitis as candidates for medical management have as a secondary benefit the identification of persons whose contacts could be at risk of becoming infected. A so-called look-back program has been recommended to identify persons who were transfused before 1992 with blood from a donor found subsequently to have hepatitis C. In addition, anti-HCV testing, once recommended for persons born between 1945 and 1965, has now been expanded to include all persons 18 year or older, independent of risk factors. Groups at higher risk and for whom testing is recommended include anyone who received a blood transfusion or a transplanted organ before the introduction of second-generation screening tests in 1992, those who ever used injection drugs (or took other illicit drugs by noninjection routes), chronically hemodialyzed patients, persons with clotting disorders who received clotting factors made before 1987 from pooled blood products, persons with elevated aminotransferase levels, health workers exposed to HCV-positive blood or contaminated needles, recipients of blood or organs from a donor found to be positive for hepatitis C, persons with HIV infection, health care and public safety

personnel following a needle stick or other nonpercutaneous exposure to HCV-infected material, sexual partners of persons with hepatitis C, and children born to HCV-positive mothers (Table 339-4).

For stable, monogamous sexual partners, sexual transmission of hepatitis C is unlikely, and sexual barrier precautions are not recommended. For persons with multiple sexual partners or with sexually transmitted diseases, the risk of sexual transmission of hepatitis C is increased, and barrier precautions (latex condoms) are recommended. A person with hepatitis C should avoid sharing such items as razors, toothbrushes, and nail clippers with sexual partners and family members. No special precautions are recommended for babies born to mothers with hepatitis C, and breast-feeding does not have to be restricted.

Hepatitis E For prevention of hepatitis E, Ig derived from HEV-endemic populations does not appear to be effective. Two safe and effective three-dose (0, 1, and 6 months), recombinant genotype 1 capsid protein vaccines, which protect against other genotypes as well, have been shown in randomized, placebo-controlled trials to be highly protective against symptomatic acute hepatitis E. A Chinese vaccine, Hecolin, achieved 100% 12-month efficacy and was licensed in China in 2011; its long-lasting protection (87% efficacy) was documented for up to 4.5 years. A second vaccine developed by GlaxoSmithKline and the U.S. Army vaccine achieved a 12-month 96% efficacy. The second vaccine was never developed commercially. The Chinese vaccine is available in China but is not FDA approved or available in the United States.

FURTHER READING

- B GJ, S BL (eds). *Eliminating the Public Health Problem of Hepatitis B and C in the United States: Phase One Report*. Washington DC, National Academies Press, 2016.
- C D C P : Recommendations for the identification of chronic hepatitis C virus infection among persons born during 1945–1965. MMWR Morb Mortal Wkly Rep 61(RR-4):1, 2012.
- C M-H et al: Long-term effects of hepatitis B immunization of infants in preventing liver cancer. Gastroenterology 151:472, 2016.
- D Y et al: Update on hepatitis E virology: Implications for clinical practice. J Hepatol 65:200, 2016.
- D MM et al: Chronic hepatitis C virus infection in the United States, National Health and Nutrition Examination Survey 2003–2010. Ann Intern Med 160:293, 2014.
- D I et al: Current epidemiology of hepatitis E virus infection in the United States: Low seroprevalence in the National Health and Nutrition Survey. Hepatology 60:815, 2014.
- D M et al: Recommendations of the Advisory Committee on Immunization Practices for use of hepatitis A vaccine for persons experiencing homelessness. MMWR Morb Mortal Wkly Rep 68:153, 2019.
- D F et al: The mechanism of HCV entry into host cells. Prog Mol Biol Transl Sci 129:63, 2015.
- E BR et al: Toward a more accurate estimate of the prevalence of hepatitis C in the United States. Hepatology 62:1353, 2015.
- E A S L : EASL clinical practice guidelines on hepatitis E virus infection. J Hepatol 68:1256, 2018.
- F M et al: Hepatitis A outbreaks associated with drug use and homelessness—California, Kentucky, Michigan, and Utah 2017. MMWR Morb Mortal Wkly Rep 67:1208, 2018.
- F M et al: Advisory Committee on Immunization Practices. Recommended adult immunization schedule, United States, 2020. Ann Intern Med 172:337, 2020.
- G D et al: Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis and liver failure on the waitlist for liver transplantation. Gastroenterology 152:1090, 2017.
- J JB et al: The spread of hepatitis C virus genotype 1a in North America: A retrospective phylogenetic study. Lancet Infect Dis 16:698, 2016.
- K C et al: Pathogenesis of and new therapies for hepatitis D. Gastroenterology 156:461, 2019.
- L MH et al: Chronic hepatitis B prevalence among foreign-born and U.S.-born adults in the United States, 1999–2016. Hepatology 71:431, 2020.
- L MH et al: Chronic hepatitis C virus infection increases mortality from hepatic and extrahepatic diseases: A community-based long-term prospective study. J Infect Dis 206:469, 2012.
- L SM et al: Type A viral hepatitis: A summary and update on the molecular virology, epidemiology, pathogenesis, and prevention. J Hepatol 68:167, 2018.
- L H-H et al: Changing hepatitis D virus epidemiology in a hepatitis B virus endemic area with a national vaccination program. Hepatology 61:1870, 2016.
- N NP et al: Update: Recommendations of the Advisory Committee on Immunization Practices for use of hepatitis A vaccine for postexposure prophylaxis and for preexposure prophylaxis for international travel. MMWR Morb Mortal Wkly Rep 67:1216, 2018.
- P CQ et al: Tenofovir to prevent hepatitis B transmission in mothers with high viral load. N Engl J Med 374:2324, 2016.
- P O C : Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: A modelling study. Lancet Gastroenterol Hepatol 3:383, 2018.
- P O HCV C : Global prevalence and genotype distribution of hepatitis C virus infection in 2015: A modelling study. Lancet Gastroenterol Hepatol 2:161, 2017.
- R M et al: Hepatitis delta: The rediscovery. Clin Liver Dis 17:475, 2013.
- R H et al: Prevalence of chronic hepatitis B virus (HBV) infection in U.S. households: National Health and Nutrition Examination Survey (NHANES), 1988–2012. Hepatology 63:388, 2016.
- R CL et al: Advisory Committee on Immunization Practices recommended immunization schedule for children and adolescents aged 18 years or younger—United States, 2020. MMWR Morb Mortal Wkly Rep 69:130, 2020.
- R ES et al: Prevalence of hepatitis C virus infection in US States and District of Columbia, 2013–2016. JAMA Network Open 1:e186371, 2018.
- R AB et al: Vital signs: Newly reported acute and chronic hepatitis C cases—United States, 2009–2018. MMWR Morb Mortal Wkly Rep 69:399, 2020.
- S S et al: Prevention of hepatitis B virus infection in the United States: Recommendation of the Advisory Committee on Immunization Practices. MMWR Morb Mortal Wkly Rep 67:1, 2018.
- S S et al: CDC recommendations for hepatitis C screening among adults—United States, 2020. MMWR Recommend Rep 69(No. RR #2):1, 2020.
- S A et al: Estimations of worldwide prevalence of chronic hepatitis B virus infection: A systematic review of data published between 1965 and 2013. Lancet 386:1546, 2015.
- S C et al: The hepatitis delta virus: replication and pathogenesis. J Hepatol 64:S102, 2016.
- T C et al: Hepatitis B virus infection. Lancet 384:2053, 2014.
- U.S. P S T F : Screening for hepatitis B virus infection in pregnant women: US Preventive Services Task Force reaffirmation recommendation statement. JAMA 322:349, 2019.
- U.S. P S T F : Screening for hepatitis C virus infection in adolescents and adults: US Preventive Services Task Force recommendation statement. JAMA 323:970, 2020.
- W I et al: Screening and treatment program to eliminate hepatitis C in Egypt. N Engl J Med 382:1166, 2020.

Liver injury is a possible consequence of ingestion of any xenobiotic, including industrial toxins, pharmacologic agents, and complementary and alternative medications (CAMs). Among patients with acute liver failure, drug-induced liver injury (DILI) is the most common cause, and evidence for hepatotoxicity detected during clinical trials for drug development is the most common reason for failure of compounds to reach approval status. DILI requires careful history-taking to identify unrecognized exposure to chemicals used in work or at home, drugs taken by prescription or bought over the counter, and herbal or dietary supplement medicines. Hepatotoxic drugs can injure the hepatocyte directly, for example, via a free-radical or metabolic intermediate that causes peroxidation of membrane lipids and that results in liver cell injury. Alternatively, a drug or its metabolite may activate components of the innate or adaptive immune system, stimulate apoptotic pathways, or initiate damage to bile excretory pathways (Fig. 340-1). Interference with bile canalicular pumps can allow endogenous bile acids, which can injure the liver, to accumulate. Such secondary injury, in turn, may lead to necrosis of hepatocytes; injure bile ducts, producing cholestasis; or block pathways of lipid movement, inhibit protein synthesis, or impair mitochondrial oxidation of fatty acids, resulting in lactic acidosis and intracellular triglyceride accumulation (expressed histologically as microvesicular steatosis). In other instances, drug metabolites sensitize hepatocytes to toxic cytokines. The differences observed between susceptible and nonsusceptible drug recipients may be attributable to human leukocyte antigen (HLA) haplotypes that determine binding of drug-related haptens on the cell surface as well as to polymorphisms in elaboration of competing, protective cytokines, as has been suggested for acetaminophen hepatotoxicity (see below). Immune mechanisms may include cytotoxic lymphocytes or antibody-mediated cellular cytotoxicity. In addition, a role has been shown for activation of nuclear transporters, such as the constitutive androstane receptor (CAR) or, more recently, the pregnane X receptor (PXR), in the induction of drug hepatotoxicity.

■ DRUG METABOLISM

Most drugs, which are water-insoluble, undergo a series of metabolic steps, culminating in a water-soluble form appropriate for renal or biliary excretion. This process begins with oxidation or methylation mediated initially by the microsomal mixed function oxygenases, cytochrome P450 (phase I reaction), followed by glucuronidation or sulfation (phase II reaction) or inactivation by glutathione. Most drug hepatotoxicity is the result of formation of a phase I toxic metabolite, but glutathione depletion, precluding inactivation of harmful compounds by glutathione S-transferase, can contribute as well by ensuring that the toxic compound is not abrogated.

■ LIVER INJURY CAUSED BY DRUGS

In general, two major types of chemical hepatotoxicity have been recognized: (1) direct toxic and (2) idiosyncratic. As shown in Table 340-1, direct toxic hepatitis occurs with predictable regularity in individuals exposed to the offending agent and is dose-dependent. The latent period between exposure and liver injury is usually short (often several hours), although clinical manifestations may be delayed for 24–48 h. Agents producing toxic hepatitis are generally systemic poisons or are converted in the liver to toxic metabolites. The direct hepatotoxins result in morphologic abnormalities that are reasonably characteristic and reproducible for each toxin. Examples of rare toxins currently include carbon tetrachloride and trichloroethylene that characteristically produce a centrilobular zonal necrosis. The hepatotoxic octapeptides of *Amanita phalloides* usually produce massive hepatic necrosis; the lethal dose of the toxin is ~10 mg, the amount found in a single

deathcap mushroom. Acetaminophen, the prime example of a direct toxin, is discussed below.

In idiosyncratic drug reactions, the occurrence of liver injury is infrequent (1 in 10^3 – 10^5 patients) and unpredictable; the response is not as clearly dose-dependent as is injury associated with direct hepatotoxins, and liver injury may occur at any time after exposure to the drug but typically between 5 and 90 days following its initiation. Although regarded as not dose-related in the fashion of direct toxins, most agents causing idiosyncratic toxicity are given at relatively high daily doses, typically exceeding 100 mg, suggesting a role for dose—drugs with low potency must be given in higher doses that engender greater chances for “off-target” effects. Likewise, drugs given in milligram amounts are of high potency and rarely cause liver or other off-target effects. Adding to the difficulty of predicting or identifying idiosyncratic drug hepatotoxicity is the occurrence of mild, transient, nonprogressive serum aminotransferase elevations that resolve with continued drug use. Such “adaptation,” the mechanism of which is unknown, is well recognized for drugs such as isoniazid (INH), valproate, phenytoin, and HMG-CoA reductase inhibitors (statins). Extrahepatic manifestations of hypersensitivity, such as rash, arthralgias, fever, leukocytosis, and eosinophilia, occur in a small fraction of patients with idiosyncratic hepatotoxic drug reactions but are characteristic for certain drugs (phenytoin, trimethoprim-sulfamethoxazole) and not others. Both primary immunologic injury and direct hepatotoxicity related to idiosyncratic differences in generation of toxic metabolites have been invoked to explain idiosyncratic drug reactions. The most current data implicate the adaptive immune system responding to the formation of immune stimulatory compounds resulting from phase I metabolic activation of the offending drug. Differences in host susceptibility may result from varying kinetics of toxic metabolite generation and genetic polymorphisms in downstream drug-metabolizing pathways or cytokine activation; in addition, certain HLA haplotypes have been associated with hepatotoxicity of certain drugs such as amoxicillin-clavulanate and flucloxacillin. Occasionally, however, the clinical features of an allergic reaction (prominent tissue eosinophilia, autoantibodies, etc.) are difficult to ignore and suggest activation of IgE pathways. A few instances of drug hepatotoxicity are observed to be associated with autoantibodies, including a class of antibodies to liver-kidney microsomes, anti-LKM2, directed against a cytochrome P450 enzyme. Four agents that specifically have a phenotype of autoimmune hepatitis with a high likelihood of positive antinuclear antibodies (ANAs) include nitrofurantoin, minocycline, hydralazine, and α -methyldopa.

Idiosyncratic reactions lead to a morphologic pattern that is more variable than those produced by direct toxins; a single agent is often capable of causing a variety of lesions, although certain patterns tend to predominate. Depending on the agent involved, idiosyncratic hepatitis may result in a clinical and morphologic picture indistinguishable from that of viral hepatitis (e.g., INH or ciprofloxacin). So-called hepatocellular injury is the most common form, featuring spotty necrosis in the liver lobule with a predominantly lymphocytic infiltrate resembling that observed in acute hepatitis A, B, or C. Drug-induced cholestasis ranges from mild to increasingly severe: (1) bland cholestasis with limited hepatocellular injury (e.g., estrogens, 17, α -substituted androgens); (2) inflammatory cholestasis (e.g., amoxicillin-clavulanic acid [the most frequently implicated antibiotic among cases of DILI], oxacillin, erythromycin estolate); (3) sclerosing cholangitis (e.g., after intrahepatic infusion of the chemotherapeutic agent floxuridine for hepatic metastases from a primary colonic carcinoma); and (4) disappearance of bile ducts, “ductopenic” cholestasis or vanishing bile duct syndrome, similar to that observed in chronic rejection (Chap. 345) following liver transplantation (e.g., carbamazepine, levofloxacin). Cholestasis may result from binding of drugs to canalicular membrane transporters, accumulation of toxic bile acids resulting from canalicular pump failure, or genetic defects in canalicular transporter proteins. Clinically, the distinction between a hepatocellular and a cholestatic reaction is indicated by the R value, the ratio of alanine aminotransferase (ALT) to alkaline phosphatase values, both expressed as multiples of the upper limit of normal. An R value of >5.0 is associated with hepatocellular

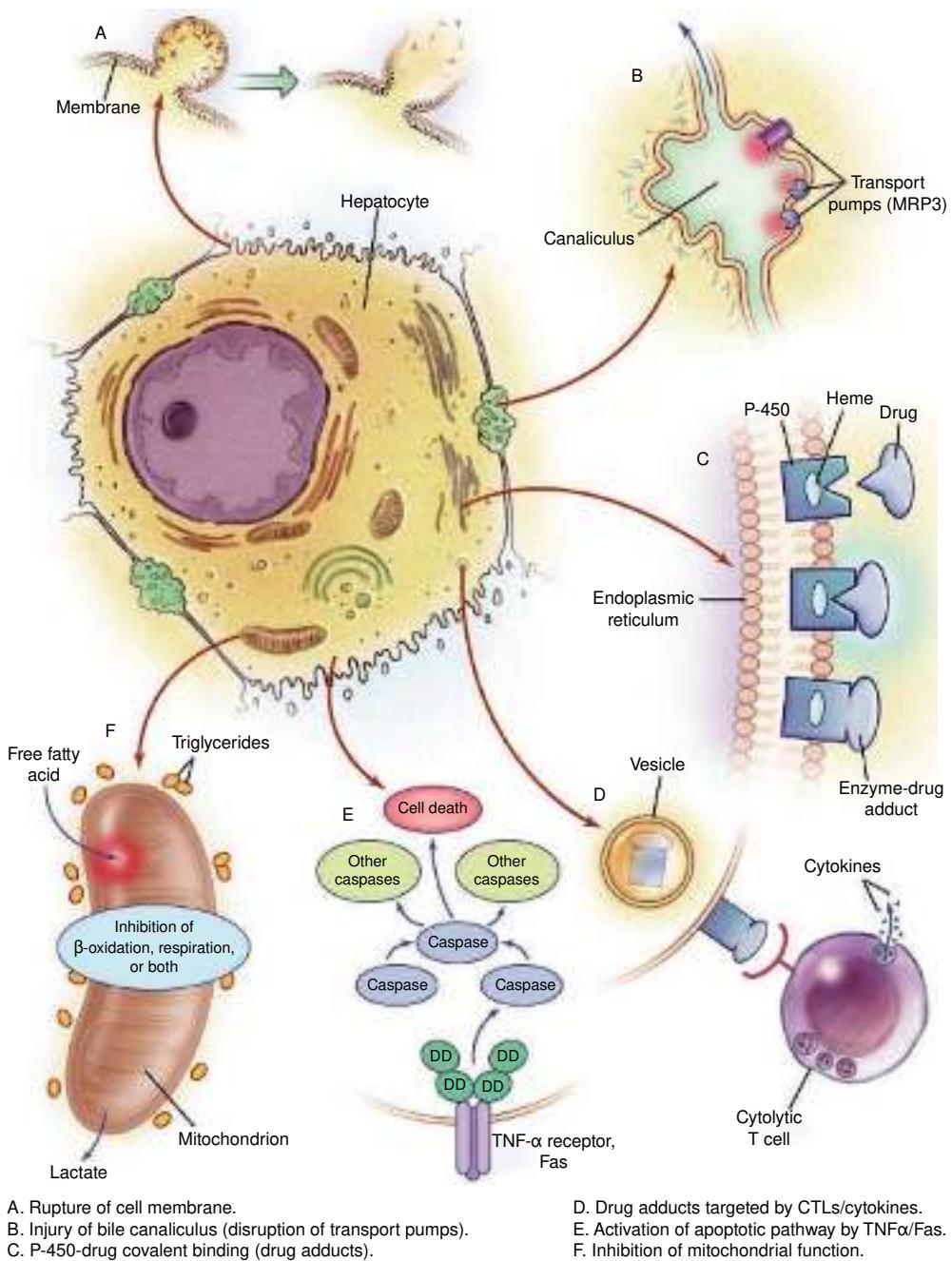


FIGURE 340-1 Potential mechanisms of drug-induced liver injury. The normal hepatocyte may be affected adversely by drugs through (A) disruption of intracellular calcium homeostasis that leads to the disassembly of actin fibrils at the surface of the hepatocyte, resulting in blebbing of the cell membrane, rupture, and cell lysis; (B) disruption of actin filaments next to the canalculus (the specialized portion of the cell responsible for bile excretion), leading to loss of villous processes and interruption of transport pumps such as multidrug resistance-associated protein 3 (MRP3), which, in turn, prevents the excretion of bilirubin and other organic compounds; (C) covalent binding of the heme-containing cytochrome P450 enzyme to the drug, thus creating nonfunctioning adducts; (D) migration of these enzyme-drug adducts to the cell surface in vesicles to serve as target immunogens for cytolytic attack by T cells, stimulating an immune response involving cytolytic T cells and cytokines; (E) activation of apoptotic pathways by tumor necrosis factor α (TNF- α) receptor or Fas (DD denotes death domain), triggering the cascade of intercellular caspases, resulting in programmed cell death; or (F) inhibition of mitochondrial function by a dual effect on both β -oxidation and the respiratory-chain enzymes, leading to failure of free fatty acid metabolism, a lack of aerobic respiration, and accumulation of lactate and reactive oxygen species (which may disrupt mitochondrial DNA). Toxic metabolites excreted in bile may damage bile-duct epithelium (not shown). CTLs, cytolytic T lymphocytes. (From WM Lee: Drug-induced hepatotoxicity. *N Engl J Med* 349:474, 2003. Copyright © 2003, Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

injury, $R < 2.0$ with cholestatic injury, and R between 2.0 and 5.0 with mixed hepatocellular-cholestatic injury.

Morphologic alterations may also include hepatic granulomas (e.g., sulfonamides) or macrovesicular or microvesicular steatosis or

steatohepatitis. Severe hepatotoxicity associated with steatohepatitis, most likely a result of mitochondrial toxicity, was recognized with certain antiretroviral therapies, although most of these drugs have been withdrawn (Chap. 202). Another potential target for idiosyncratic drug

TABLE 340-1 Some Features of Toxic and Drug-Induced Hepatic Injury

FEATURES	DIRECT TOXIC EFFECT ^a		IDIOSYNCRATIC ^b			OTHER ^c ESTROGENS/ ANDROGENIC STEROIDS
	CARBON TETRACHLORIDE	ACETAMINOPHEN	AMOXICILLIN- CLAVULANATE	ISONIAZID	CIPROFLOXACIN	
Predictable and dose-related toxicity	+	+	0	0	0	+
Latent period	Short	Short	Delayed onset	Variable	May be short	Variable
Arthralgia, fever, rash, eosinophilia	0	0	0	0	0	0
Liver morphology	Necrosis, fatty infiltration	Centrilobular necrosis	Mixed hepatocellular/cholestatic	Hepatocellular injury resembling viral hepatitis	Hepatocellular injury resembling viral hepatitis	Cholestasis without portal inflammation

^aThe drugs listed are typical examples.

hepatotoxicity is sinusoidal lining cells; when these are injured, such as by high-dose chemotherapeutic agents (e.g., cyclophosphamide, melphalan, busulfan) administered prior to bone marrow transplantation, veno-occlusive disease can result. Nodular regenerative hyperplasia, a subtle form of portal hypertension, may also result from vascular injury to portal or hepatic venous endothelium following systemic chemotherapy, such as with oxaliplatin, as part of adjuvant treatment for colon cancer.

Not all adverse hepatic drug reactions can be classified as either toxic or idiosyncratic. For example, oral contraceptives, which combine estrogenic and progestational compounds, may result in impairment of liver tests and, occasionally, jaundice; however, they do not produce necrosis or fatty change, manifestations of hypersensitivity are generally absent, and susceptibility to the development of oral contraceptive-induced cholestasis appears to be genetically determined. Such estrogen-induced cholestasis is more common in women with cholestasis of pregnancy, a disorder linked to genetic defects in multidrug resistance-associated canalicular transporter proteins.

Any idiosyncratic reaction that occurs in <1:10,000 recipients will go unrecognized in most clinical trials, which involve at most several thousand subjects. The U.S. Food and Drug Administration (FDA) and pharmaceutical companies have learned to look for even subtle indications of serious toxicity and monitor regularly the number of trial subjects in whom any aminotransferase elevations develop, as a possible surrogate for more serious toxicity. Even more valid as a predictor of severe hepatotoxicity is the occurrence of jaundice in patients enrolled in a clinical drug trial, so-called "Hy's Law," named after Dr. Hyman Zimmerman, one of the pioneers of the field of drug hepatotoxicity. He recognized that, if jaundice occurred during a phase 3 trial, more serious liver injury was likely, with a 10:1 ratio between cases of jaundice and liver failure (i.e., 10 patients with jaundice would result in 1 patient with acute liver failure). Thus, the finding of such Hy's Law (jaundiced) cases during drug development often portends failure of approval, particularly if any of the subjects sustains a bad outcome. Troglitazone, a peroxisome proliferator-activated receptor γ agonist, was the first in its class of thiazolidinedione insulin-sensitizing agents. Although in retrospect, Hy's Law cases of jaundice had occurred during phase 3 trials, no instances of liver failure were recognized until well after the drug was introduced, emphasizing the importance of postmarketing surveillance in identifying toxic drugs and in leading to their withdrawal from use. Fortunately, such hepatotoxicity is not characteristic of the second-generation thiazolidinediones rosiglitazone and pioglitazone; in clinical trials, the frequency of aminotransferase elevations in patients treated with these medications did not differ from that in placebo recipients, and isolated reports of liver injury among recipients are extremely rare. Since troglitazone was withdrawn from the market in 2001, no fully approved drugs have had to be withdrawn from the market by the FDA. Several agents have received black box warnings indicating that caution is needed; overall, the industry and FDA in concert have been able to avert severe toxicity in approved agents over the past 20 years.

Proving that an episode of liver injury is caused by a drug (causality) is difficult in many cases. DILI is nearly always a presumptive diagnosis, and many other disorders produce a similar clinicopathologic picture. Thus, causality may be difficult to establish and requires

several separate supportive assessment variables to lead to a high level of certainty, including temporal association (time of onset, time to resolution), clinical-biochemical features, type of injury (hepatocellular vs cholestatic), extrahepatic features, likelihood that a given agent is to blame based on its past record, and exclusion of other potential causes. Scoring systems such as the Roussel-Uclaf Causality Assessment Method (RUCAM) yield residual uncertainty and have not been adopted widely. Currently, the U.S. Drug-Induced Liver Injury Network (DILIN) relies on a structured expert opinion process requiring detailed data on each case and a comprehensive review by three experts who arrive at a consensus on a five-degree scale of likelihood (definite, highly likely, probable, possible, unlikely); however, this approach is not practical for routine clinical application.

Generally, drug hepatotoxicity is not more frequent in persons with underlying chronic liver disease, although the severity of the outcome may be amplified. Reported exceptions include hepatotoxicity of aspirin, methotrexate, INH (only in certain experiences), antiretroviral therapy for HIV infection, and certain drugs such as conditioning regimens for bone marrow transplantation in the presence of hepatitis C.

TREATMENT

Toxic and Drug-Induced Hepatic Disease

Treatment is largely supportive, except in acetaminophen hepatotoxicity (for which *N*-acetylcysteine is effective, see below). Acute liver failure develops in 10% of patients with DILI; spontaneous recovery, once that threshold is reached, occurs in <30%, and liver transplantation is performed in >40% of those who reach the level of severity of acute liver failure (coagulopathy and hepatic encephalopathy) (Chap. 345). Withdrawal of the suspected agent is indicated at the first sign of an adverse reaction or when aminotransferase levels reach five times the upper limit of normal. A number of studies have suggested that lethal outcomes follow continued use of an agent in the face of symptoms and signs of liver injury. In the case of the direct toxins, liver involvement should not divert attention from renal or other organ involvement, which may also threaten survival. Agents used occasionally but of questionable value include glucocorticoids for DILI with allergic features, silybin for mushroom poisoning, and ursodeoxycholic acid for cholestatic drug hepatotoxicity; these medications have been shown to be effective and cannot be recommended. A double-blind, randomized controlled trial of the use of *N*-acetylcysteine for nonacetaminophen acute liver failure, including cases of DILI, demonstrated benefit, particularly for patients with early-stage hepatic encephalopathy; however, the drug has not been approved by FDA for this indication.

In Table 340-2, several classes of chemical agents are listed together with examples of the pattern of liver injury they produce. Certain drugs appear to be responsible for the development of chronic as well as acute hepatic injury. For example, nitrofurantoin, minocycline, hydralazine, and methyldopa have been associated with moderate to severe chronic hepatitis with autoimmune features. Methotrexate, tamoxifen, and

TABLE 340-2 Principal Alterations of Hepatic Morphology Produced by Some Commonly Used Drugs and Chemicals^a

PRINCIPAL MORPHOLOGIC CHANGE	CLASS OF AGENT	EXAMPLE
Cholestasis	Anabolic steroid	Methyl testosterone, many other body-building supplements
	Antibiotic	Erythromycin estolate, nitrofurantoin, rifampin, amoxicillin-clavulanic acid, oxacillin
	Anticonvulsant	Carbamazepine
	Antidepressant	Duloxetine, mirtazapine, tricyclic antidepressants
	Anti-inflammatory	Sulindac
	Antiplatelet	Clopidogrel
	Antihypertensive	Irbesartan, fosinopril
	Antithyroid	Methimazole
	Calcium channel blocker	Nifedipine, verapamil
	Immunosuppressive	Cyclosporine
	Lipid-lowering	Ezetimibe
	Oncotherapeutic	Anabolic steroids, busulfan, tamoxifen, irinotecan, cytarabine, temozolomide
	Oral contraceptive	Norethynodrel with mestranol
	Oral hypoglycemic	Chlorpropamide
	Tranquilizer	Chlorpromazine ^b
Fatty liver	Antiarrhythmic	Amiodarone
	Antibiotic	Tetracycline (high-dose, IV)
	Anticonvulsant	Valproic acid
	Antiviral	Dideoxynucleosides (e.g., zidovudine), protease inhibitors (e.g., indinavir, ritonavir)
	Oncotherapeutic	Asparaginase, methotrexate, tamoxifen
Hepatitis	Anesthetic	Halothane, fluothane
	Antiandrogen	Flutamide
	Antibiotic	Isoniazid, ^c rifampicin, nitrofurantoin, telithromycin, minocycline, ^d pyrazinamide, trovafloxacin ^e
	Anticonvulsant	Phenytoin, carbamazepine, valproic acid, phenobarbital
	Antidepressant	Iproniazid, amitriptyline, trazodone, venlafaxine, fluoxetine, paroxetine, duloxetine, sertraline, nefazodone ^f
	Antifungal	Ketoconazole, fluconazole, itraconazole
	Antihypertensive	Methyldopa, ^c captopril, enalapril, lisinopril, losartan
	Anti-inflammatory	Ibuprofen, indomethecin, diclofenac, sulindac, bromfenac
	Antipsychotic	Risperidone
	Antiviral	Zidovudine, didanosine, stavudine, nevirapine, ritonavir, indinavir, tipranavir, zalcitabine
	Calcium channel blocker	Nifedipine, verapamil, diltiazem
	Cholinesterase inhibitor	Tacrine
	Diuretic	Chlorothiazide
	Laxative	Oxyphenisatin ^{c,e}
	Norepinephrine reuptake inhibitor	Atomoxetine
	Oral hypoglycemic	Troglitazone, ^g acarbose
Mixed hepatitis/cholestatic	Antibiotic	Amoxicillin-clavulanic acid, trimethoprim-sulfamethoxazole
	Antibacterial	Clindamycin
	Antifungal	Terbinafine
	Antihistamine	Cyproheptadine
	Immunosuppressive	Azathioprine
	Lipid-lowering	Nicotinic acid, lovastatin, ezetimibe
Toxic (necrosis)	Analgesic	Acetaminophen
	Hydrocarbon	Carbon tetrachloride
	Metal	Yellow phosphorus
	Mushroom	<i>Amanita phalloides</i>
	Solvent	Dimethylformamide
Granulomas	Antiarrhythmic	Quinidine, diltiazem
	Antibiotic	Sulfonamides
	Anticonvulsant	Carbamazepine
	Anti-inflammatory	Phenylbutazone
	Xanthine oxidase inhibitor	Allopurinol
Vascular injury	Chemotherapeutic	Oxaliplatin, melphalan

^aSeveral agents cause more than one type of liver lesion and appear under more than one category. ^bRarely associated with primary biliary cirrhosis-like lesion.^cOccasionally associated with chronic hepatitis or bridging hepatic necrosis or cirrhosis. ^dAssociated with an autoimmune hepatitis-like syndrome. ^eWithdrawn from use because of severe hepatotoxicity.

amiodarone have been implicated in the development of cirrhosis. Portal hypertension in the absence of cirrhosis, termed *nodular regenerative hyperplasia*, may result from alterations in hepatic architecture produced by excessive intake of vitamin A or following chemotherapy with oxaliplatin. Oral contraceptives have been implicated in the development of focal nodular hyperplasia or hepatic adenoma (both benign lesions) and, rarely, hepatocellular carcinoma and hepatic vein occlusion (Budd-Chiari syndrome). Another unusual lesion, peliosis hepatitis (blood cysts of the liver), has been observed in some patients treated with anabolic or contraceptive steroids. The existence of these hepatic disorders expands the spectrum of liver injury induced by chemical agents and emphasizes the need for a thorough drug history in all patients with liver dysfunction. The comprehensive, authoritative LiverTox website, which contains up-to-date information on DILI, is available as a valuable reference through the National Institutes of Health and the National Library of Medicine (livertox.nih.gov).

The following are patterns of adverse hepatic reactions for some prototypic agents.

■ ACETAMINOPHEN HEPATOTOXICITY DIRECT TOXIN

Acetaminophen represents the most prevalent cause of acute liver failure in the Western world; up to 72% of patients with acetaminophen hepatotoxicity in Scandinavia—somewhat lower frequencies in the United Kingdom and the United States—progress to encephalopathy and coagulopathy. Acetaminophen causes dose-related centrilobular hepatic necrosis after single-time-point ingestions, as intentional self-harm, or over extended periods, as unintentional overdoses, when multiple drug preparations or inappropriate drug amounts are used daily for several days, for example, for relief of pain or fever. In these instances, 8 g/d, twice the daily recommended maximum dose, over several days can readily lead to liver failure. Use of opioid-acetaminophen combinations appears to be particularly harmful, because habituation to the opioid may occur with a gradual increase in opioid-acetaminophen combination dosing over days or weeks. A single dose of 10–15 g, occasionally less, may produce clinical evidence of liver injury. Fatal fulminant disease is usually (although not invariably) associated with ingestion of ≥ 25 g. Blood levels of acetaminophen correlate with severity of hepatic injury (levels >300 $\mu\text{g}/\text{mL}$ 4 h after ingestion are predictive of the development of severe damage; levels <150 $\mu\text{g}/\text{mL}$ suggest that hepatic injury is highly unlikely). Nausea, vomiting, diarrhea, abdominal pain, and shock are early manifestations occurring 4–12 h after ingestion. Then 24–48 h later, when these features are abating, hepatic injury becomes apparent. Maximal abnormalities and hepatic failure are evident 3–5 days after ingestion, and aminotransferase levels exceeding 10,000 IU/L are not uncommon (i.e., levels far exceeding those in patients with viral hepatitis). Renal failure and myocardial injury may be present. Whether or not a clear history of overdose can be elicited, clinical suspicion of acetaminophen hepatotoxicity should be raised by the presence of the extremely high aminotransferase levels in association with low bilirubin levels that are characteristic of this hyperacute injury. This biochemical signature should trigger further questioning of the subject if possible; however, outright denial (or denial of high doses) or altered mentation may confound diagnostic efforts. In this setting, a presumptive diagnosis is reasonable, and the proven antidote, *N*-acetylcysteine, is both safe and will be effective if given early (within 12 h) but is also used even when injury has evolved.

Acetaminophen is metabolized predominantly by a phase II reaction to innocuous sulfate and glucuronide metabolites; however, a small proportion is metabolized by a phase I reaction to a hepatotoxic metabolite formed from the parent compound by cytochrome P450 CYP2E1. This metabolite, *N*-acetyl-p-benzoquinone-imine (NAPQI), is detoxified by binding to “hepatoprotective” glutathione to become harmless, water-soluble mercapturic acid, which undergoes renal excretion. When excessive amounts of NAPQI are formed, or when glutathione levels are low, glutathione levels are depleted and overwhelmed, permitting covalent binding to nucleophilic hepatocyte macromolecules forming acetaminophen-protein “adducts.” These

adducts, which can be measured in serum by high-performance liquid chromatography, hold promise as diagnostic markers of acetaminophen hepatotoxicity, and a point-of-care assay for acetaminophen-Cys adducts is under development. The binding of acetaminophen to hepatocyte macromolecules is believed to lead to hepatocyte necrosis; the precise sequence and mechanism are unknown. Hepatic injury may be potentiated by prior administration of alcohol, phenobarbital, INH, or other drugs; by conditions that stimulate the mixed-function oxidase system; or by conditions such as starvation (including inability to maintain oral intake during severe febrile illnesses) that reduce hepatic glutathione levels. Alcohol induces cytochrome P450 CYP2E1; consequently, increased levels of the toxic metabolite NAPQI may be produced in chronic alcoholics after acetaminophen ingestion, but the role of alcohol in potentiating acute acetaminophen injury is still debated. Alcohol also suppresses hepatic glutathione production. Therefore, in chronic alcoholics, the toxic dose of acetaminophen may be as low as 2 g, and alcoholic patients should be warned specifically about the dangers of even standard doses of this commonly used drug. In a 2006 study, aminotransferase elevations were identified in 31–44% of normal subjects treated for 14 days with the maximal recommended dose of acetaminophen, 4 g daily (administered alone or as part of an acetaminophen-opioid combination); because these changes were transient and never associated with bilirubin elevation, the clinical relevance of these findings remains to be determined. Although underlying hepatitis C virus (HCV) infection was found to be associated with an increased risk of acute liver injury in patients hospitalized for acetaminophen overdose, generally, in patients with nonalcoholic liver disease, acetaminophen taken in recommended doses is well tolerated. Acetaminophen use in cirrhotic patients has not been associated with hepatic decompensation. On the other hand, because of the link between acetaminophen use and liver injury and because of the limited safety margin between safe and toxic doses, the FDA has recommended that the daily dose of acetaminophen be reduced from 4 g to 3 g (even lower for persons with chronic alcohol use), that all acetaminophen-containing products be labeled prominently as containing acetaminophen, and that the potential for liver injury be prominent in the packaging of acetaminophen and acetaminophen-containing products. Within opioid combination products, the limit for the acetaminophen component has been lowered to 325 mg per tablet.

TREATMENT

Acetaminophen Overdosage

Treatment includes gastric lavage, supportive measures, and oral administration of activated charcoal or cholestyramine to prevent absorption of residual drug. Neither charcoal nor cholestyramine appears to be effective if given >30 min after acetaminophen ingestion; if they are used, the stomach lavage should be done before other agents are administered orally. The chances of possible, probable, and high-risk hepatotoxicity can be derived from a nomogram plot (Fig. 340-2), readily available in emergency departments, as a function of measuring acetaminophen plasma levels 4–8 h after ingestion. In patients with high acetaminophen blood levels (>200 $\mu\text{g}/\text{mL}$ measured at 4 h or >100 $\mu\text{g}/\text{mL}$ at 8 h after ingestion), the administration of *N*-acetylcysteine reduces markedly the severity of hepatic necrosis. This agent provides sulfhydryl donor groups to replete glutathione, which is required to render harmless toxic metabolites that would otherwise bind covalently via sulfhydryl linkages to cell proteins, resulting in the formation of drug metabolite-protein adducts. Therapy should be begun within 8 h of ingestion but may be at least partially effective when given as late as 24–36 h after overdose. Routine use of *N*-acetylcysteine has substantially reduced the occurrence of fatal acetaminophen hepatotoxicity. *N*-acetylcysteine may be given orally but is more commonly used as an IV solution, with a loading dose of 140 mg/kg over 1 h, followed by 70 mg/kg every 4 h for 15–20 doses. Whenever a patient with potential acetaminophen hepatotoxicity is encountered, a local poison control center should be contacted. Treatment can be

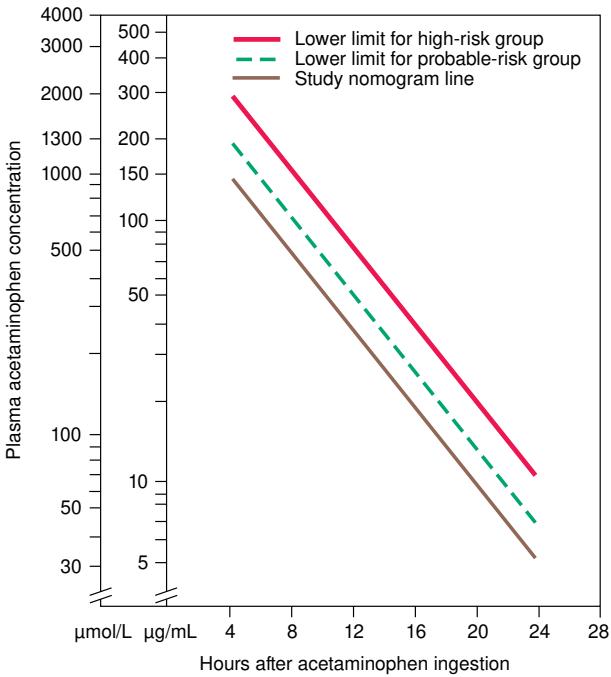


FIGURE 340-2 Nomogram to define risk of acetaminophen hepatotoxicity according to initial plasma acetaminophen concentration. (Reproduced with permission from *Pediatrics*, 55:871. Copyright © 1975 by the AAP.)

stopped when plasma acetaminophen levels indicate that the risk of liver damage is low. If signs of hepatic failure (e.g., progressive jaundice, coagulopathy, confusion) occur despite *N*-acetylcysteine therapy for acetaminophen hepatotoxicity, liver transplantation may be the only option. Early arterial blood lactate levels among such patients with acute liver failure may distinguish patients highly likely to require liver transplantation (lactate levels >3.5 mmol/L) from those likely to survive without liver replacement. Acute renal injury occurs in nearly 75% of patients with severe acetaminophen injury but is virtually always self-limited.

Survivors of acute acetaminophen overdose rarely, if ever, have ongoing liver injury or sequela but may be subject to repeat overdosing.

■ ISONIAZID HEPATOTOXICITY TOXIC AND IDIOSYNCRATIC REACTION

INH remains central to most antituberculous prophylactic and therapeutic regimens, despite its long-standing recognition as a hepatotoxin. In 10% of patients treated with INH, elevated serum aminotransferase levels develop during the first few weeks of therapy; however, these elevations in most cases are self-limited, are mild (values for ALT <200 IU/L), and resolve despite continued drug use. This adaptive response allows continuation of the agent if symptoms and progressive enzyme elevations do not follow the initial elevations. Acute hepatocellular DILI secondary to INH is evident with a variable latency period up to 6 months and is more frequent in alcoholics and patients taking certain other medications, such as barbiturates, rifampin, and pyrazinamide. If the clinical threshold of encephalopathy is reached, severe hepatic injury is likely to be fatal or to require liver transplantation. Liver biopsy reveals morphologic changes similar to those of viral hepatitis or bridging hepatic necrosis. Substantial liver injury appears to be age-related, increasing substantially after age 35; the highest frequency is in patients over age 50, and the lowest is in patients under the age of 20. Even for patients >50 years of age monitored carefully during therapy, hepatotoxicity occurs in only ~2%, well below the risk estimate derived from earlier experiences. Fever, rash, eosinophilia, and other manifestations of drug allergy are distinctly unusual. Antibodies

to INH have been detected in INH recipients, but a link to causality of liver injury remains unclear. A clinical picture resembling chronic hepatitis has been observed in a few patients. Many public health programs that require INH prophylaxis for a positive tuberculin skin test or blood test (Quantiferon or T-Spot) include monthly monitoring of aminotransferase levels, although this practice has been called into question. Even more effective in limiting serious outcomes may be encouraging patients to be alert for symptoms such as nausea, fatigue, or jaundice, because most fatalities occur in the setting of continued INH use despite clinically apparent illness. The incidence of severe INH toxicity may be declining as a result of less frequent use and/or better management.

■ SODIUM VALPROATE HEPATOTOXICITY TOXIC AND IDIOSYNCRATIC REACTION

Sodium valproate, an anticonvulsant useful in the treatment of petit mal and other seizure disorders, has been associated with the development of severe hepatic toxicity and, rarely, fatalities, predominantly in children but also in adults. Among children listed as candidates for liver transplantation, valproate is the most common antiepileptic drug implicated. Asymptomatic elevations of serum aminotransferase levels have been recognized in as many as 45% of treated patients. These “adaptive” changes, however, appear to have no clinical importance, because major hepatotoxicity is not seen in the majority of patients despite continuation of drug therapy. In the rare patients in whom jaundice, encephalopathy, and evidence of hepatic failure are found, examination of liver tissue reveals microvesicular fat and bridging hepatic necrosis, predominantly in the centrilobular zone. Bile duct injury may also be apparent. Most likely, sodium valproate is not directly hepatotoxic, but its metabolite, 4-pentenoic acid, may be responsible for hepatic injury. Valproate hepatotoxicity is more common in persons with mitochondrial enzyme deficiencies and may be ameliorated by IV administration of carnitine, which valproate therapy can deplete. Valproate toxicity has been linked to HLA haplotypes (*DR4* and *B'1502*) and to mutations in mitochondrial DNA polymerase gamma 1.

■ NITROFURANTOIN HEPATOTOXICITY IDIOSYNCRATIC REACTION

This commonly used antibiotic for urinary tract infections may cause an acute hepatitis leading to fatal outcome or, more frequently, chronic hepatitis of varying severity but indistinguishable from autoimmune hepatitis. These two scenarios may reflect the frequent use and reuse of the drug for treatment of recurrent cystitis in women. Although most toxic agents manifest injury within 6 months of first ingestion, nitrofurantoin may have a longer latency period, in part perhaps because of its intermittent, recurrent use. Autoantibodies to nuclear components, smooth muscle, and mitochondria are seen and may subside after resolution of injury; however, glucocorticoid or other immunosuppressive medication may be necessary to resolve the autoimmune injury, and cirrhosis may be seen in cases that are not recognized quickly. Interstitial pulmonary fibrosis presenting as chronic cough and dyspnea may be present and resolve slowly with medication withdrawal. Histologic findings are identical to those of autoimmune hepatitis. A similar disease pattern can be observed with minocycline, which is used repeatedly for the treatment of acne in teenagers, as well as with hydralazine and α -methylldopa.

■ AMOXICILLIN CLAVULANATE HEPATOTOXICITY IDIOSYNCRATIC MIXED REACTION

Currently, the most common agent implicated as causing DILI in the United States and in Europe is amoxicillin-clavulanate (most frequent brand name: Augmentin). This medication causes a very specific syndrome of mixed or primarily cholestatic injury. Because hepatotoxicity may follow amoxicillin-clavulanate therapy after a relatively long latency period, the liver injury may begin to manifest after the drug has been withdrawn. The high prevalence of hepatotoxicity reflects in part the very frequent use of this drug for respiratory tract infections, including community-acquired pneumonia. The mechanism of

hepatotoxicity is unclear, but the liver injury is thought to be caused by amoxicillin toxicity that is potentiated in some way by clavulanate, which itself appears not to be toxic. Symptoms include nausea, anorexia, fatigue, and jaundice—which may be prolonged—with pruritus. Rash is quite uncommon. On occasion, amoxicillin-clavulanate, like other cholestatic hepatotoxic drugs, causes permanent injury to small bile ducts, leading to the so-called “vanishing bile duct syndrome.” In vanishing bile duct syndrome, initially, liver injury is minimal except for severe cholestasis; however, over time, histologic evidence of bile duct abnormalities is replaced by a paucity and eventual absence of discernible ducts on subsequent biopsies.

■ AMIODARONE HEPATOTOXICITY TOXIC AND IDIOSYNCRATIC REACTION

Therapy with this potent antiarrhythmic drug is accompanied in 15–50% of patients by modest elevations of serum aminotransferase levels that may remain stable or diminish despite continuation of the drug. Such abnormalities may appear days to many months after beginning therapy. A proportion of those with elevated aminotransferase levels have detectable hepatomegaly, and clinically important liver disease develops in <5% of patients. Features that represent a direct effect of the drug on the liver and that are common to the majority of long-term recipients are ultrastructural phospholipidosis, unaccompanied by clinical liver disease, and interference with hepatic mixed-function oxidase metabolism of other drugs. The cationic amphiphilic drug and its major metabolite desethylamiodarone accumulate in hepatocyte lysosomes and mitochondria and in bile duct epithelium. The relatively common elevations in aminotransferase levels are also considered a predictable, dose-dependent, direct hepatotoxic effect. On the other hand, in the rare patient with clinically apparent, symptomatic liver disease, liver injury resembling that seen in alcoholic liver disease is observed. The so-called pseudoalcoholic liver injury can range from steatosis, to alcoholic hepatitis-like neutrophilic infiltration and Mallory's hyaline, to cirrhosis. Electron-microscopic demonstration of phospholipid-laden lysosomal lamellar bodies can help to distinguish amiodarone hepatotoxicity from typical alcoholic hepatitis. This category of liver injury appears to be a metabolic idiosyncrasy that allows hepatotoxic metabolites to be generated. Rarely, an acute idiosyncratic hepatocellular injury resembling viral hepatitis or cholestatic hepatitis occurs. Hepatic granulomas have occasionally been observed. Because amiodarone has a long half-life, liver injury may persist for months after the drug is stopped.

■ ANABOLIC STEROIDS CHOLESTATIC REACTION

The most common form of liver injury caused by CAMs is the profound cholestasis associated with anabolic steroids used by body builders. Unregulated agents sold in gyms and health food stores as diet supplements, which are taken by athletes to improve their performance, may contain anabolic steroids. In a young male, jaundice that is accompanied by a cholestatic, rather than a hepatic, laboratory profile almost invariably will turn out to be caused by the use of one of a variety of androgen congeners. Such agents have the potential to injure bile transport pumps and to cause intense cholestasis; the time to onset is variable, and resolution, which is the rule, may require many weeks to months. Initially, anorexia, nausea, and malaise may occur, followed by pruritus in some but not all patients. Serum aminotransferase levels are usually <100 IU/L, and serum alkaline phosphatase levels are generally moderately elevated with bilirubin levels frequently exceeding 342 µmol/L (20 mg/dL). Examination of liver tissue reveals cholestasis without substantial inflammation or necrosis. Anabolic steroids have also been used by prescription to treat bone marrow failure. In this setting, hepatic centrilobular sinusoidal dilatation and peliosis hepatis have been reported in rare patients, as have hepatic adenomas and hepatocellular carcinoma. Recently, a large series of cases with a uniform phenotype has been described. Unfortunately, no genomic signature has become evident despite the unique features of the injury. No permanent sequelae are evident besides prolonged jaundice, lasting frequently 10 weeks or more.

■ TRIMETHOPRIM SULFAMETHOXAZOLE HEPATOTOXICITY IDIOSYNCRATIC REACTION

This antibiotic combination is used routinely for urinary tract infections in immunocompetent persons and for prophylaxis against and therapy of *Pneumocystis jirovecii* pneumonia in immunosuppressed persons (transplant recipients, patients with AIDS). With its increasing use, its occasional hepatotoxicity is being recognized with growing frequency. Its likelihood is unpredictable, but when it occurs, trimethoprim-sulfamethoxazole hepatotoxicity follows a relatively uniform latency period of several weeks and is often accompanied by eosinophilia, rash, and other features of a hypersensitivity reaction, including the drug reaction with eosinophilia and systemic symptoms (DRESS) syndrome. Biochemically and histologically, acute hepatocellular necrosis predominates, but cholestatic features are quite frequent. Occasionally, cholestasis without necrosis occurs, and very rarely, a severe cholangiolitic pattern of liver injury is observed. In most cases, liver injury is self-limited, but rare fatalities have been recorded. The hepatotoxicity is attributable to the sulfamethoxazole component of the drug and is similar in features to that seen with other sulfonamides; tissue eosinophilia and granulomas may be seen. The risk of trimethoprim-sulfamethoxazole hepatotoxicity is increased in persons with HIV infection. In a recent study, unique HLA associations in European Americans and in African Americans have been identified.

■ HMG COA REDUCTASE INHIBITORS STATINS IDIOSYNCRATIC MIXED HEPATOCELLULAR AND CHOLESTATIC REACTION

Between 1 and 2% of patients taking lovastatin, simvastatin, pravastatin, fluvastatin, or one of the newer statin drugs for the treatment of hypercholesterolemia experience asymptomatic, reversible elevations (greater than threefold) of aminotransferase activity. Acute hepatitis-like histologic changes, centrilobular necrosis, and centrilobular cholestasis have been described in a very small number of cases. In a larger proportion, minor aminotransferase elevations appear during the first several weeks of therapy. Careful laboratory monitoring can distinguish between patients with minor, transitory changes, who may continue therapy, and those with more profound and sustained abnormalities, who should discontinue therapy. Because clinically meaningful aminotransferase elevations are so rare after statin use and do not differ in meta-analyses from the frequency of such laboratory abnormalities in placebo recipients, a panel of liver experts recommended to the National Lipid Association's Safety Task Force that liver test monitoring was not necessary in patients treated with statins and that statin therapy need not be discontinued in patients found to have asymptomatic isolated aminotransferase elevations during therapy. Statin hepatotoxicity is not increased in patients with chronic hepatitis C, hepatic steatosis, or other underlying liver diseases, and statins can be used safely in these patients.

■ ALTERNATIVE AND COMPLEMENTARY MEDICINES IDIOSYNCRATIC HEPATITIS, STEATOSIS

Herbal medications that are of scientifically unproven efficacy and that lack prospective safety oversight by regulatory agencies account currently for >20% of DILI in the United States. Besides anabolic steroids, the most common category of dietary or herbal products is weight loss agents. Included among the herbal remedies associated with toxic hepatitis are Jin Bu Huan, xiao-chai-hu-tang, germander, chaparral, senna, mistletoe, skullcap, gentian, comfrey (containing pyrrolizidine alkaloids), ma huang, bee pollen, valerian root, pennyroyal oil, kava, celadine, Impila (*Callilepis laureola*), LipoKinetix, Hydroxycut, OxyElite Pro, Herbalife, herbal nutritional supplements, and herbal teas containing *Camellia sinensis* (green tea extract). Well characterized are the acute hepatitis-like histologic lesions following Jin Bu Huan use: focal hepatocellular necrosis, mixed mononuclear portal tract infiltration, coagulative necrosis, apoptotic hepatocyte degeneration, tissue eosinophilia, and microvesicular steatosis. Megadoses of vitamin A can injure the liver, as can pyrrolizidine alkaloids, which often contaminate Chinese herbal preparations and can cause a

veno-occlusive injury leading to sinusoidal hepatic vein obstruction. Because some alternative medicines induce toxicity via active metabolites, alcohol and drugs that stimulate cytochrome P450 enzymes may enhance the toxicity of some of these products. Conversely, some alternative medicines also stimulate cytochrome P450 and may result in or amplify the toxicity of recognized drug hepatotoxins. In many instances, herbal and dietary supplements actually contain chemicals rather than only leaves, roots, and bark. Antirheumatic "herbs" have been found to contain a nonsteroidal anti-inflammatory drug (NSAID) such as diclofenac, for example. Given the widespread use of such poorly defined herbal preparations, hepatotoxicity is likely to be encountered with increasing frequency; therefore, a drug history in patients with acute and chronic liver disease should include use of "alternative medicines" and other nonprescription preparations sold in so-called health food stores.

CHECKPOINT INHIBITOR AND OTHER IMMUNOTHERAPIES FOR CANCER

The introduction of a new class of immunotherapeutic agents for melanoma and other cancers has ushered in a new kind of hepatotoxicity, that associated with activation of the immune response. The three classes of immune-active molecules are cytotoxic T-lymphocyte-associated antigen-4 (CTLA-4), programmed cell death receptor 1 (PD-1), and programmed cell death receptor ligand 1 (PD-L1). Within weeks of beginning treatment with any one of several agents, including ipilimumab (CTLA-4), pembrolizumab (PD-1), or nivolumab (PD-1), an active hepatitis evolves that is associated with positive ANAs and appears to respond to glucocorticoid therapy. Histologically, liver histology does not resemble autoimmune hepatitis but, instead, a nonspecific hepatic injury, assumed to result from the release of host modulation of anti-self-immune responses. Immune-mediated injury to thyroid, muscle, and colon is also commonly seen. Few deaths have been reported related to these immunotherapies; while these novel agents may need to be halted temporarily, in many cases, they can be restarted (and are tolerated better on retreatment) if patients are showing a favorable antitumor response.

HIGHLY ACTIVE ANTIRETROVIRAL THERAPY FOR HIV INFECTION MITOCHONDRIAL TOXIC, IDIOSYNCRATIC, STEATOSIS; HEPATOCELLULAR, CHOLESTATIC, AND MIXED

The recognition of drug hepatotoxicity in persons with HIV infection is complicated in this population by the many alternative causes of liver injury (chronic viral hepatitis, fatty infiltration, infiltrative disorders, mycobacterial infection, etc.), but drug hepatotoxicity associated with highly active antiretroviral therapy (HAART) was a common type of liver injury in HIV-infected persons in the early days of HIV therapy; however, it is less frequent now (*Chap. 202*). Implicated most frequently are combinations including the nucleoside analogue reverse transcriptase inhibitors zidovudine, didanosine, and, to a lesser extent, stavudine; the protease inhibitors ritonavir and indinavir (and amprenavir when used together with ritonavir), as well as tipranavir; and the nonnucleoside reverse transcriptase inhibitors nevirapine and, to a lesser extent, efavirenz. Distinguishing the impact of HAART hepatotoxicity in patients with HIV and hepatitis virus co-infection is made challenging by the following: (1) both chronic hepatitis B and hepatitis C can affect the natural history of HIV infection and the response to HAART, and (2) HAART can have an impact on chronic viral hepatitis. For example, immunologic reconstitution with HAART can result in immunologically mediated liver-cell injury in patients with chronic hepatitis B co-infection if treatment with an antiviral agent for hepatitis B (e.g., nucleoside analogues such as tenofovir) is withdrawn. Infection with HIV, especially with low CD4+ T-cell counts, has been reported to increase the rate of hepatic fibrosis associated with chronic hepatitis C, and HAART therapy can increase levels of serum aminotransferases and HCV RNA in patients with hepatitis C co-infection. Didanosine or stavudine should not be used with ribavirin in patients with HIV/HCV co-infection because of an increased risk of severe mitochondrial toxicity and lactic acidosis.

A

Kurt J Isselbacher, MD, contributed to this chapter in previous editions of *Harrison's*.

FURTHER READING

- A J et al: Sclerosing cholangitis-like changes on magnetic resonance cholangiography in patients with drug-induced liver injury. *Clin Gastroenterol Hepatol* 17:789, 2019.
- B ES, H JL: Categorization of drugs implicated in causing liver injury: Critical assessment based upon published case reports. *Hepatology* 63:590, 2016.
- C N et al: Features and outcomes of 899 patients with drug-induced liver injury: The DILIN prospective study. *Gastroenterology* 148:1340, 2015.
- C ET et al: A missense variant in PTPN22 is a risk factor for drug-induced liver injury. *Gastroenterology* 156:1707, 2019.
- B YS et al: Features of autoimmune hepatitis in patients with drug-induced liver injury. *Clin Gastroenterol Hepatol* 15:103, 2017.
- K N, D LD (eds): *Drug-Induced Liver Disease*, 3rd ed. London, Elsevier/Academic Press, 2013.
- K DE: Histopathological challenges in suspected drug-induced liver injury. *Liver Int* 38:198, 2018.
- L WM et al: Intravenous N-acetylcysteine improves transplant-free survival in early stage non-acetaminophen acute liver failure. *Gastroenterology* 137:856, 2009.
- P TB et al: Hepatotoxicity from immune checkpoint inhibitors: A systematic review and management recommendation. *Hepatology* 72:315, 2020.
- S A et al: Severe and protracted cholestasis in 44 young men taking bodybuilding supplements: Assessment of genetic, clinical and chemical risk factors. *Aliment Pharmacol Ther* 49:1195, 2019.

341

Chronic Hepatitis

Jules L. Dienstag



Chronic hepatitis represents a series of liver disorders of varying causes and severity in which hepatic inflammation and necrosis continue for at least 6 months. Milder forms are nonprogressive or only slowly progressive, while more severe forms may be associated with scarring and architectural reorganization, which, when advanced, lead ultimately to cirrhosis. Several categories of chronic hepatitis have been recognized. These include chronic viral hepatitis, drug-induced chronic hepatitis (*Chap. 340*), and autoimmune chronic hepatitis. In many cases, clinical and laboratory features are insufficient to allow assignment into one of these three categories; these "idiopathic" cases are also believed to represent autoimmune chronic hepatitis. Finally, clinical and laboratory features of chronic hepatitis are observed occasionally in patients with such hereditary/metabolic disorders as Wilson's disease (copper overload), α_1 antitrypsin deficiency (*Chaps. 344 and 415*), and nonalcoholic fatty liver disease (*Chap. 343*) and even occasionally in patients with alcoholic liver injury (*Chap. 342*). Although all types of chronic hepatitis share certain clinical, laboratory, and histopathologic features, chronic viral and chronic autoimmune hepatitis are sufficiently distinct to merit separate discussions. **For discussion of acute hepatitis, see *Chap. 339*.**

CLASSIFICATION OF CHRONIC HEPATITIS

Common to all forms of chronic hepatitis are histopathologic distinctions based on localization and extent of liver injury. These vary from the milder forms, previously labeled *chronic persistent hepatitis* and *chronic lobular hepatitis*, to the more severe form, formerly called *chronic active hepatitis*. When first defined, these

designations were believed to have prognostic implications, which were not corroborated by subsequent observations. Categorization of chronic hepatitis based primarily on histopathologic features has been replaced by a more informative classification based on a combination of clinical, serologic, and histologic variables. Classification of chronic hepatitis is based on (1) its *cause*; (2) its histologic activity, or *grade*; and (3) its degree of progression based on level of fibrosis, or *stage*. Thus, neither clinical features alone nor histologic features—requiring liver biopsy or noninvasive markers of fibrosis—alone are sufficient to characterize and distinguish among the several categories of chronic hepatitis.

■ CLASSIFICATION BY CAUSE

Clinical and serologic features allow the establishment of a diagnosis of *chronic viral hepatitis*, caused by hepatitis B, hepatitis B plus D, or hepatitis C; *autoimmune hepatitis*, including several subcategories, I and II, based on serologic distinctions; *drug-associated chronic hepatitis*; and a category of unknown cause, or *cryptogenic chronic hepatitis* (Table 341-1). These are addressed in more detail below.

■ CLASSIFICATION BY GRADE

Grade, a histologic assessment of necroinflammatory activity, is based on examination of the liver biopsy. An assessment of important histologic features includes the degree of *periportal necrosis* and the disruption of the limiting plate of periportal hepatocytes by inflammatory cells (so-called *piecemeal necrosis* or *interface hepatitis*); the degree of confluent necrosis that links or forms bridges between vascular structures—between portal tract and portal tract or even more important bridges between portal tract and central vein—referred to as *bridging necrosis*; the degree of hepatocyte degeneration and focal necrosis within the lobule; and the degree of *portal inflammation*. Several scoring systems that take these histologic features into account have been devised, and the most popular are the histologic activity index (HAI), used commonly in the United States, and the METAVIR score, used in Europe (Table 341-2). Based on the presence and degree of these features of histologic activity, chronic hepatitis can be graded as mild, moderate, or severe.

■ CLASSIFICATION BY STAGE

The stage of chronic hepatitis, which reflects the level of progression of the disease, is based on the degree of hepatic fibrosis. When fibrosis is so extensive that fibrous septa surround parenchymal nodules and alter the normal architecture of the liver lobule, the histologic lesion is defined as *cirrhosis*. Staging is based on the degree of fibrosis as categorized on a numerical scale 0–6 (HAI) or 0–4 (METAVIR) (Table 341-2). Several

noninvasive approaches have been introduced to provide approximations of hepatic histologic stage, including serum biomarkers of fibrosis; fibrosis scores such as FIB-4, a validated algorithm based on such routine lab tests as aspartate and alanine aminotransferase (AST and ALT) levels and platelet counts (PLT) ($\text{age} [\text{years}] \times \text{AST}/\text{PLT} \times \text{ALT}^{1/2}$); and imaging determinations of liver elasticity.

CHRONIC VIRAL HEPATITIS

Both of the enterically transmitted forms of viral hepatitis, hepatitis A and E, are self-limited and do not cause chronic hepatitis (rare reports notwithstanding in which acute hepatitis A serves as a trigger for the onset of autoimmune hepatitis in genetically susceptible patients or in which hepatitis E [Chap. 339] can cause chronic liver disease in immunosuppressed hosts, for example, after liver transplantation). In contrast, the entire clinicopathologic spectrum of chronic hepatitis occurs in patients with chronic viral hepatitis B and C as well as in patients with chronic hepatitis D superimposed on chronic hepatitis B.

■ CHRONIC HEPATITIS B

The likelihood of chronicity after acute hepatitis B varies as a function of age. Infection at birth is associated with clinically silent acute infection but a 90% chance of chronic infection, whereas infection in young adulthood in immunocompetent persons is typically associated with clinically apparent acute hepatitis but a risk of chronicity of only ~1%. Most cases of chronic hepatitis B among adults, however, are recognized in patients who never had a recognized episode of clinically apparent acute viral hepatitis. The degree of liver injury (grade) in patients with chronic hepatitis B is variable, ranging from none in inactive carriers to mild to moderate to severe. Among adults with chronic hepatitis B, histologic features are of prognostic importance. In one long-term study of patients with chronic hepatitis B, investigators found a 5-year survival rate of 97% for patients with mild chronic hepatitis, 86% for patients with moderate to severe chronic hepatitis, and only 55% for patients with chronic hepatitis and postnecrotic cirrhosis. The 15-year survival in these cohorts was 77%, 66%, and 40%, respectively. On the other hand, more recent observations do not allow us to be so sanguine about the prognosis in patients with mild chronic hepatitis; among such patients followed for 1–13 years, progression to more severe chronic hepatitis and cirrhosis has been observed in more than a quarter of cases.

More important to consider than histology alone in patients with chronic hepatitis B is the degree of hepatitis B virus (HBV) replication. As reviewed in Chap. 339, chronic HBV infection can occur in the presence or absence of serum hepatitis B e antigen (HBeAg), and

TABLE 341-1 Clinical and Laboratory Features of Chronic Hepatitis

TYPE OF HEPATITIS	DIAGNOSTIC TEST(S)	AUTOANTIBODIES	THERAPY
Chronic hepatitis B	HBsAg, IgG anti-HBc, HBeAg, HBV DNA	Uncommon	IFN- α , PEG IFN- α Oral agents: First-line: entecavir, tenofovir Second-line: lamivudine, adefovir, telbivudine
Chronic hepatitis C	Anti-HCV, HCV RNA	Anti-LKM 1 ^a	PEG IFN- α plus ribavirin ^b Direct-acting oral agents: sofosbuvir, ledipasvir, velpatasvir ritonavir-boosted paritaprevir, ombitasvir, dasabuvir, elbasvir, grazoprevir, daclatasvir, simeprevir
Chronic hepatitis D	Anti-HDV, HDV RNA, HBsAg, IgG anti-HBc	Anti-LKM 3	IFN- α , PEG IFN- α ^c
Autoimmune hepatitis	ANA ^d (homogeneous), anti-LKM 1 (\pm) Hyperglobulinemia	ANA, anti-LKM 1 anti-SLA ^e	Prednisone, azathioprine
Drug-associated	—	Uncommon	Withdraw drug
Cryptogenic	All negative	None	Prednisone (?), azathioprine (?)

^aAntibodies to liver-kidney microsomes type 1 (autoimmune hepatitis type II and some cases of hepatitis C). ^bSupplanted in almost all cases by combinations of the direct-acting antiviral agents listed (see www.hcvguidelines.org). ^cEarly clinical trials suggested benefit of IFN- α therapy; PEG IFN- α is as effective, if not more so, and has supplanted standard IFN- α . ^dAntinuclear antibody (autoimmune hepatitis type I). ^eAntibodies to soluble liver antigen (autoimmune hepatitis type III).

Abbreviations: HBc, hepatitis B core; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; HCV, hepatitis C virus; HDV, hepatitis D virus; IFN- α , interferon α ; IgG, immunoglobulin G; LKM, liver-kidney microsome; PEG IFN- α , pegylated interferon α ; SLA, soluble liver antigen.

TABLE 341-2 Histologic Grading and Staging of Chronic Hepatitis

		HISTOLOGIC ACTIVITY INDEX (HAI) ^a	METAVIR ^b		
HISTOLOGIC FEATURE		SEVERITY	SCORE	SEVERITY	SCORE
Necroinflammatory Activity (grade)					
Periportal necrosis, including piecemeal necrosis and/or bridging necrosis (BN)	None	0	None	0	
	Mild	1	Mild	1	
	Mild/moderate	2	Moderate	2	
	Moderate	3	Severe	3	
	Severe	4	Bridging necrosis	Yes	
Intralobular necrosis	Confluent	—None	0	None or mild	0
		—Focal	1	Moderate	1
		—Zone 3 some	2	Severe	2
		—Zone 3 most	3		
		—Zone 3 + BN few	4		
		—Zone 3 + BN multiple	5		
		—Panacinar/multiacinar	6		
Focal	—None	0			
	—≤1 focus/10x field	1			
	—2–4 foci/10x field	2			
	—5–10 foci/10x field	3			
	—>10 foci/10x field	4			
Portal Inflammation		None	0		
		Mild	1		
		Moderate	2		
		Moderate/markd	3		
		Marked	4		
		Total	0–18	A0–A3 ^c	
Fibrosis (stage)					
None			0	F0	
			1	F1	
			2	F1	
			3	F2	
			4	F3	
			5	F4	
		Total	6	F4	
Portal fibrosis—some			6	4	
Portal fibrosis—most					
Bridging fibrosis—few					
Bridging fibrosis—many					
Incomplete cirrhosis					
Cirrhosis					

^aIshak K, Baptista A, Bianchi L, et al: Histologic grading and staging of chronic hepatitis. J Hepatol 22:696, 1995.

^bBedossa P, Poynard T, French METAVIR Cooperative Study Group: An algorithm for grading activity in chronic hepatitis C. Hepatology 24:289, 1996. ^cNecroinflammatory grade: A0 = none; A1 = mild; A2 = moderate; A3 = severe.

generally, for both HBeAg-reactive and HBeAg-negative chronic hepatitis B, the level of HBV DNA correlates with the level of liver injury and risk of progression. In *HBeAg-reactive chronic hepatitis B*, two phases have been recognized based on the relative level of HBV replication. The relatively *replicative phase* is characterized by the presence in the serum of HBeAg and HBV DNA levels well in excess of 10^3 – 10^4 IU/mL, sometimes exceeding 10^9 IU/mL; by the presence in the liver of detectable intrahepatocyte nucleocapsid antigens (primarily hepatitis B core antigen [HBcAg]); by high infectivity; and by accompanying liver injury. In contrast, the relatively *nonreplicative phase* is characterized by the absence of the conventional serum marker of HBV replication (HBeAg), the appearance of anti-HBe, levels of HBV DNA below a threshold of $\sim 10^3$ IU/mL, the absence of intrahepatocytic HBcAg, limited infectivity, and minimal liver injury. Patients in the relatively replicative phase tend to have more severe chronic hepatitis, whereas those in the relatively nonreplicative phase tend to have minimal or mild chronic hepatitis or to be inactive hepatitis B carriers. The likelihood in a patient with HBeAg-reactive chronic hepatitis B of converting spontaneously from relatively replicative to nonreplicative infection is $\sim 10\%$ per year. Distinctions in HBV replication and in histologic category, however, do not always coincide. In patients with HBeAg-reactive chronic HBV infection, especially when acquired at birth or in

40–55 years, older than that for HBeAg-reactive chronic hepatitis B; these mutations prevent translation of HBeAg from the precore component of the HBV genome (precore mutants) or are characterized by downregulated transcription of precore mRNA (core-promoter mutants; Chap. 339). Although their levels of HBV DNA tend to be lower than among patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B can have progressive liver injury (complicated by cirrhosis and HCC) and experience episodic reactivation of liver disease reflected in fluctuating levels of aminotransferase activity ("flares"). The biochemical and histologic activity of HBeAg-negative disease tends to correlate closely with levels of HBV replication, unlike the case mentioned above of Asian patients with HBeAg-reactive chronic hepatitis B during the early decades of their HBV infection. Worth reiterating, the level of HBV replication is the most important risk factor for the ultimate development of cirrhosis and HCC in both HBeAg-reactive (beyond the early decades of "relatively nonreplicative" infection) and HBeAg-negative patients. Although levels of HBV DNA are lower and more readily suppressed by therapy to undetectable levels in HBeAg-negative (compared to HBeAg-reactive) chronic hepatitis B, achieving sustained responses that permit discontinuation of antiviral therapy is less likely in HBeAg-negative patients (see below). Inactive carriers are patients

early childhood, as recognized commonly in Asian countries, a dichotomy is common between very high levels of HBV replication during the early decades of life (when the level of apparent host immunologic tolerance of HBV is relatively high) and negligible levels of liver injury; during this phase of chronic hepatitis B, the level of viral replication does not correlate with liver injury or late complications. Yet despite the relatively immediate, apparently benign nature of liver disease for many decades in this population, in the middle decades, activation of liver injury emerges as what appears to be the relative tolerance of the host to HBV declines, and these patients with childhood-acquired HBV infection are ultimately at increased risk later in life for cirrhosis, hepatocellular carcinoma (HCC) (Chap. 82), and liver-related death; the link between high-level HBV replication and these late liver complications has been demonstrated convincingly in, and confined mostly to, persons in their middle decades, especially age ≥ 40 . A discussion of the pathogenesis of liver injury in patients with chronic hepatitis B appears in Chap. 339.

HBeAg-negative chronic hepatitis B (i.e., chronic HBV infection with active virus replication and readily detectable HBV DNA but without HBeAg [anti-HBe-reactive]) is more common than HBeAg-reactive chronic hepatitis B in Mediterranean and European countries and in Asia (and, correspondingly, in HBV genotypes other than A). Compared to patients with HBeAg-reactive chronic hepatitis B, patients with HBeAg-negative chronic hepatitis B have HBV DNA levels several orders of magnitude lower (usually no more than 10^2 – 10^6 IU/mL) than those observed in the HBeAg-reactive subset. Most such cases represent precore or core-promoter mutations acquired late in the natural history of the disease (mostly early-life onset; age range

with circulating hepatitis B surface antigen (HBsAg), normal serum aminotransferase levels, minimal or no histologic evidence of liver injury, undetectable HBeAg, and levels of HBV DNA that are either undetectable or present at a threshold of $\leq 10^3$ IU/mL. This serologic profile occurs not only in inactive carriers but also in patients with HBeAg-negative chronic hepatitis B during periods of relative inactivity; distinguishing between the two requires sequential biochemical and virologic monitoring over many months.

The spectrum of *clinical features* of chronic hepatitis B is broad, ranging from asymptomatic infection to debilitating disease or even end-stage, fatal hepatic failure. As noted above, the onset of the disease tends to be insidious in most patients, apart from the very few in whom chronic disease follows failure of resolution of clinically apparent acute hepatitis B. **The clinical and laboratory features associated with progression from acute to chronic hepatitis B are discussed in Chap. 339.**

Fatigue is a common symptom, and persistent or intermittent *jaundice* is a common feature in severe or advanced cases. Intermittent deepening of jaundice and recurrence of malaise and anorexia, as well as worsening fatigue, are reminiscent of acute hepatitis; such exacerbations may occur spontaneously, often coinciding with evidence of virologic reactivation; may lead to progressive liver injury; and, when superimposed on well-established cirrhosis, may cause hepatic decompensation. Complications of cirrhosis occur in end-stage chronic hepatitis and include ascites, edema, bleeding gastroesophageal varices, hepatic encephalopathy, coagulopathy, and hypersplenism. Occasionally, these complications bring the patient to initial clinical attention. Extrahepatic complications of chronic hepatitis B, similar to those seen during the prodromal phase of acute hepatitis B, are associated with tissue deposition of circulating hepatitis B antigen–antibody immune complexes. These include arthralgias and arthritis, which are common, and the rarer purpuric cutaneous lesions (leukocytoclastic vasculitis), immune-complex glomerulonephritis, and generalized vasculitis (polyarteritis nodosa) (Chap. 363).

Laboratory features of chronic hepatitis B do not distinguish adequately between histologically mild and severe hepatitis. Aminotransferase elevations tend to be modest for chronic hepatitis B but may fluctuate in the range of 100–1000 units. As is true for acute viral hepatitis B, ALT tends to be more elevated than AST; however, once cirrhosis is established, AST tends to exceed ALT. Levels of alkaline phosphatase activity tend to be normal or only marginally elevated. In severe cases, moderate elevations in serum bilirubin (51.3–171 $\mu\text{mol/L}$ [3–10 mg/dL]) occur. Hypoalbuminemia and prolongation of the prothrombin time occur in severe or end-stage cases. Hyperglobulinemia and detectable circulating auto-antibodies are distinctly absent in chronic hepatitis B (in contrast to autoimmune hepatitis). **Viral markers of chronic HBV infection are discussed in Chap. 339.**

TREATMENT

Chronic Hepatitis B

Although progression to cirrhosis is more likely in severe than in mild or moderate chronic hepatitis B, all forms of chronic hepatitis B can be progressive, and progression occurs primarily in patients with active HBV replication. Moreover, in populations of patients with chronic hepatitis B who are at risk for HCC (Chap. 82), the risk is highest for those with continued, high-level HBV replication and lower for persons in whom initially high-level HBV DNA falls spontaneously over time. Therefore, management of chronic hepatitis B is directed at suppressing the level of virus replication. Although clinical trials tend to focus on clinical endpoints achieved over 1–2 years (e.g., suppression of HBV DNA to undetectable levels, loss of HBeAg/HBsAg, improvement in histology, normalization of ALT), these short-term gains translate into reductions in the risk of clinical progression, hepatic decompensation, HCC, liver transplantation, and death; regression of cirrhosis and of esophageal varices has been documented to follow long-term

pharmacologic suppression of HBV replication. In addition, restoration of impaired HBV-specific T-cell function has been shown following successful suppression of HBV replication with antiviral therapy. To date, eight drugs have been approved for treatment of chronic hepatitis B: injectable interferon (IFN) α and pegylated interferon (long-acting IFN bound to polyethylene glycol, PEG [PEG IFN]) and the oral agents lamivudine, adefovir dipivoxil, entecavir, telbivudine, tenofovir disoproxil fumarate (TDF), and tenofovir alafenamide (TAF).

Antiviral therapy for hepatitis B has evolved rapidly since the mid-1990s, as has the sensitivity of tests for HBV DNA. When IFN and the first oral antiviral lamivudine were evaluated in clinical trials, HBV DNA was measured by insensitive hybridization assays with detection thresholds of 10^5 – 10^6 virions/mL; when subsequent treatments were studied in clinical trials, HBV DNA was measured by sensitive amplification assays (polymerase chain reaction [PCR]) with detection thresholds of 10^1 – 10^3 viral copies/mL or IU/mL. Recognition of these distinctions is helpful when comparing results of clinical trials that established the efficacy of these therapies (reviewed below in chronologic order of publication of these efficacy trials). Of the eight approved treatments, PEG IFN, entecavir, and the two tenofovir preparations (TDF and TAF) are recommended as first-line agents, and generally, the oral agents are favored over injectable PEG IFN.

INTERFERON

IFN- α was the first approved therapy (1992) for chronic hepatitis B. Although it is no longer used to treat hepatitis B, standard IFN is important historically, having provided important lessons about antiviral therapy in general. For immunocompetent adults with *HBeAg-reactive chronic hepatitis B* (who tend to have high-level HBV DNA [$>10^5$ – 10^6 virions/mL] and histologic evidence of chronic hepatitis on liver biopsy), a 16-week course of subcutaneous IFN, 5 million units daily or 10 million units thrice weekly, resulted in a loss of HBeAg and hybridization-detectable HBV DNA (i.e., a reduction to levels below 10^5 – 10^6 virions/mL) in ~30% of patients, with a concomitant improvement in liver histology. Seroconversion from HBeAg to anti-HBe occurred in ~20%, and, in early trials, ~8% lost HBsAg. Successful IFN therapy and seroconversion were often accompanied by an acute hepatitis-like elevation in aminotransferase activity, postulated to result from enhanced cytolytic T-cell clearance of HBV-infected hepatocytes. Relapse after successful therapy was rare (1 or 2%). Responsiveness to IFN was higher in patients with low-level HBV DNA and substantial ALT elevations. Therapy with IFN was not effective in immunosuppressed persons, persons with neonatal acquisition of infection and minimal-to-mild ALT elevations, or patients with decompensated chronic hepatitis B (in whom such therapy was actually detrimental, sometimes precipitating decompensation, often associated with severe adverse effects). After HBeAg loss during IFN therapy, 80% experienced eventual loss of HBsAg and ALT normalization over the ensuing decade. In addition, improved long-term and complication-free survival as well as a reduction in the frequency of HCC were documented among IFN responders, supporting the conclusion that successful antiviral therapy improves the natural history of chronic hepatitis B.

Brief-duration IFN therapy in patients with *HBeAg-negative chronic hepatitis B* was disappointing, suppressing HBV replication transiently during therapy but almost never resulting in sustained antiviral responses; however, more protracted courses for up to 1.5 years resulted in sustained virologic/biochemical remissions documented to last for several years in ~20%.

Complications of IFN therapy include systemic “flu-like” symptoms; marrow suppression; emotional lability (irritability, depression, anxiety); autoimmune reactions (especially autoimmune thyroiditis); and miscellaneous side effects such as alopecia, rashes, diarrhea, and numbness and tingling of the extremities. With the possible exception of autoimmune thyroiditis, all these side effects are reversible upon dose lowering or cessation of therapy.

Although no longer competitive with the newer generation of antivirals, IFN did represent the first successful antiviral approach, set a standard against which to measure efficacy of subsequent drugs, and demonstrated the benefit of antiviral therapy on the natural history of chronic hepatitis B. Standard IFN has been supplanted by long-acting PEG IFN (see below), and IFN nonresponders are now treated with one of the newer oral nucleoside analogues.

LAMIVUDINE

The first of the nucleoside analogues to be approved (in 1998) for hepatitis B, the dideoxynucleoside lamivudine inhibits reverse

transcriptase activity of both HIV and HBV and is an effective agent for chronic hepatitis B; however, it is now superseded by newer, more potent, less resistance-prone agents. For a summary of its virologic, serologic, biochemical, and histologic efficacy, as well as its resistance profile, please refer to Table 341-3. In clinical trials, lamivudine therapy at daily doses of 100 mg for 48–52 weeks suppressed HBV DNA, as measured by sensitive PCR amplification assays, by a median of $\sim 5.5 \log_{10}$ copies/mL in HBeAg-positive chronic hepatitis B and $\sim 4.5 \log_{10}$ copies/mL in HBeAg-negative chronic hepatitis B (baseline HBV DNA levels

TABLE 341-3 Comparison of Pegylated Interferon (PEG IFN), Lamivudine, Adefovir, Entecavir, Telbivudine, and Tenofovir Therapy for Chronic Hepatitis B^a

FEATURE	PEG IFN ^b	LAMIVUDINE	ADEFOVIR	ENTECAVIR	TELBIVUDINE	TENOFOVIR (TDF)	TENOFOVIR (TAF)
Route of administration	Subcutaneous injection (180 µg/week)	Oral (100 mg/d)	Oral (10 mg/d)	Oral (0.5 mg/d)	Oral (600 mg/d)	Oral (300 mg/d)	Oral 25 mg/d
Status	First-line	No longer preferred	No longer preferred	First-line	No longer preferred, withdrawn	First-line	First-line
Duration of therapy ^c	48–52 weeks	≥52 weeks	≥48 weeks	≥48 weeks	≥52 weeks	≥48 weeks	48 weeks
Tolerability	Poorly tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated	Well tolerated	Well tolerated; creatinine monitoring recommended	Well tolerated
HBeAg seroconversion 1 yr Rx >1 yr Rx	18–20% NA	16–21% up to 50% at 5 yrs	12% 43% at 3 yrs ^d	21% 31% at 2 yrs 44% at 6 yrs	22% 30% at 2 yrs	21% 40% at 5 yrs	10% (14% HBeAg loss) 18% at yr 2 (HBeAg loss 22%)
Log ₁₀ HBV DNA reduction (mean copies/mL) HBeAg-reactive HBeAg-negative	4.5 4.1	5.5 4.4–4.7	Median 3.5–5 Median 3.5–3.9	6.9 5.0	6.4 5.2	6.2 4.6	Not reported in clinical trials, likely same as TDF
HBV DNA PCR negative (at then current PCR sensitivity ^e) at end of yr 1 HBeAg-reactive HBeAg-negative	10–25% 63%	36–44% 60–73%	13–21% 48–77%	67% (91% at 4 yrs) 90%	60% 88%	76% 93%	64% 94%
ALT normalization at end of yr 1 HBeAg-reactive HBeAg-negative	39% 34–38%	41–75% 62–79%	48–61% 48–77%	68% 78%	77% 74%	68% 76%	72% 83%
HBsAg loss, yr 1 >yr 1	3–4% 12% 5 yr after 1 yr of Rx	≤1% No data	0% 5% at yr 5	2% 6% at yr 6	<1% No data	3% 8% at yr 5	1% 1%
Histologic improvement (≥2 point reduction in HAI) at yr 1 HBeAg-reactive HBeAg-negative							Not included in clinical trials
Viral resistance	None	15–30% at 1 yr 70% at 5 yrs	None at 1 yr 29% at 5 yrs	≤1% at 1 yr ^g 1.2% at 6 yrs ^g	Up to 5% at yr 1 Up to 22% at yr 2	0% at yr 1 0% through yr 8	0% at yr 1 0% through yr 2
Pregnancy category	C	C	C	C	B	B	B

^aGenerally, these comparisons are based on data on each drug tested individually versus placebo in registration clinical trials; with rare exception, these comparisons are not based on head-to-head testing of these drugs. In addition, the sensitivity of HBV DNA assays increased in sensitivity over the two decades between the introduction of the earliest and latest of these approved drugs. Therefore, relative advantages and disadvantages should be interpreted cautiously. ^bAlthough standard interferon α administered daily or three times a week is approved as therapy for chronic hepatitis B, it has been supplanted by PEG IFN, which is administered once a week and is more effective. Standard interferon has no advantages over PEG IFN. ^cDuration of therapy in clinical efficacy trials; use in clinical practice may vary. ^dBecause of a computer-generated randomization error that resulted in misallocation of drug versus placebo during the second year of clinical trial treatment, the frequency of HBeAg seroconversion beyond the first year is an estimate (Kaplan-Meier analysis) based on the small subset in whom adefovir was administered correctly. ^e7% during a year of therapy (43% at year 4) in lamivudine-resistant patients. ^fDespite its category C designation, lamivudine has an extensive pregnancy safety record in women with HIV/AIDS. Abbreviations: ALT, alanine aminotransferase; HAI, histologic activity index; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; NA, not applicable; PEG IFN, pegylated interferon; PCR, polymerase chain reaction; Rx, therapy; yr, year; TAF, tenofovir alafenamide; TDF, tenofovir disoproxil fumarate.

are lower in HBeAg-negative than in HBeAg-positive chronic hepatitis B) and to undetectable levels in ~40% and ~70%, respectively. Lamivudine, which was shown to improve histology, retard hepatic fibrosis, and prevent progression to cirrhosis, was effective in patients resistant to IFN (e.g., those with high-level HBV DNA) or who had failed prior IFN therapy. As was true for IFN therapy of chronic hepatitis B, lamivudine-associated HBeAg seroconversion occurred in ~20%; patients with near-normal ALT activity tended not to experience HBeAg responses (despite suppression of HBV DNA), while patients with ALT levels \geq the upper limit of normal could expect 1-year HBeAg seroconversion rates of 50–60%. Generally, HBeAg seroconversions were confined to patients who achieved suppression of HBV DNA to $<10^4$ copies/mL (equivalent to $\sim 10^3$ IU/mL). Lamivudine-associated HBeAg responses were accompanied by a delayed posttreatment HBsAg seroconversion rate comparable to that seen after IFN. Among Western patients who experienced HBeAg responses during a year-long course of therapy and in whom the response was sustained for 4–6 months after cessation of therapy, the response was durable thereafter in the vast majority (>80%); therefore, the achievement of an HBeAg response represented a viable stopping point in therapy. Reduced durability, however, was reported in Asian patients; therefore, to support the durability of HBeAg responses, a period of consolidation therapy of ≥ 6 months in Western patients and ≥ 1 year in Asian patients was recommended after HBeAg seroconversion (see treatment guidelines below; a full 12-month consolidation period is recommended currently for treatment extension after oral-agent-induced HBeAg seroconversion). Close posttreatment monitoring was recommended to identify HBV reactivation promptly and to resume therapy. If HBeAg was unaffected by lamivudine therapy, lamivudine was continued until an HBeAg response occurred, but long-term therapy was required to suppress HBV replication and, in turn, limit liver injury; HBeAg seroconversions increased to a level of 50% after 5 years of therapy. After a cumulative course of 3 years of lamivudine therapy, necroinflammatory activity was reduced in the majority of patients, and even cirrhosis was shown to regress to precirrhotic stages in as many as three-quarters of patients.

Losses of HBsAg were few during the first year of lamivudine therapy, and this observation had been cited as an advantage of IFN-based therapy over lamivudine therapy; however, in head-to-head comparisons between standard IFN and lamivudine monotherapy, HBsAg losses were rare in both groups. Trials in which lamivudine and IFN were administered in combination failed to show a benefit of combination therapy over lamivudine monotherapy for either treatment-naïve patients or prior IFN nonresponders.

Patients with *HBeAg-negative chronic hepatitis B* (i.e., in those with precore and core-promoter HBV mutations and who lack HBeAg) cannot achieve an HBeAg response to nucleoside analogue therapy—a stopping point in HBeAg-reactive patients; almost invariably, when therapy was discontinued, reactivation was the rule. Therefore, these patients required long-term lamivudine therapy.

Clinical and laboratory side effects of lamivudine were negligible and indistinguishable from those observed in placebo recipients; however, lamivudine doses were reduced in patients with reduced creatinine clearance. During lamivudine therapy, transient ALT elevations, resembling those seen during IFN therapy and during spontaneous HBeAg-to-anti-HBe seroconversions, occurred in one-fourth of patients. These ALT elevations may result from restored cytolytic T-cell activation permitted by suppression of HBV replication. Similar ALT elevations, however, occurred at an identical frequency in placebo recipients; however, ALT elevations associated temporally with HBeAg seroconversion in clinical trials were confined to lamivudine-treated patients. When therapy was stopped after a year of therapy, two- to threefold ALT elevations occurred in 20–30% of lamivudine-treated patients, representing renewed liver-cell injury as HBV replication returned. Although these posttreatment flares were almost always transient and mild, rare severe exacerbations, especially in cirrhotic patients, were observed, mandating close and careful clinical and virologic

monitoring after discontinuation of treatment. Many authorities cautioned against discontinuing therapy in patients with cirrhosis, in whom posttreatment flares could precipitate decompensation.

Long-term monotherapy with lamivudine was associated with methionine-to-valine (M204V) or methionine-to-isoleucine (M204I) mutations, primarily at amino acid 204 in the tyrosine-methionine-aspartate-aspartate (YMDD) motif of the C domain of HBV DNA polymerase, analogous to mutations that occur in HIV-infected patients treated with this drug. During a year of therapy, YMDD mutations occurred in 15–30% of patients; the frequency increased with each year of therapy, reaching 70% at year 5. Ultimately, patients with YMDD mutants experienced degradation of clinical, biochemical, and histologic responses; therefore, if treatment was begun with lamivudine monotherapy, the emergence of lamivudine resistance, reflected clinically by a breakthrough from suppressed levels of HBV DNA and ALT, was managed by adding another antiviral to which YMDD variants are sensitive (e.g., adefovir, tenofovir; see below).

Currently, lamivudine has been eclipsed by more potent antivirals that have superior resistance profiles (see below); it is no longer recommended as first-line therapy. Still, as the first successful oral antiviral agent for use in hepatitis B, lamivudine provided proof of principle that polymerase inhibitors can achieve virologic, serologic, biochemical, and histologic benefits, including retardation and reversal of fibrosis and even of cirrhosis. In addition, lamivudine was shown to be effective in the treatment of patients with decompensated hepatitis B (for whom IFN is contraindicated), in some of whom decompensation can be reversed. Moreover, among patients with cirrhosis or advanced fibrosis, lamivudine was shown to be effective in reducing the risk of progression to hepatic decompensation and, based on subsequent population studies, the risk of HCC. In the half decade following the introduction in the United States of lamivudine therapy for hepatitis B, referral of patients with HBV-associated end-stage liver disease for liver transplantation fell by ~30%, supporting further the beneficial impact of oral antiviral therapy on the natural history of chronic hepatitis B.

Because lamivudine monotherapy in persons with HIV infection can result universally in the rapid emergence of YMDD variants, testing for HIV infection was recommended for all patients with chronic hepatitis B prior to lamivudine therapy; if HIV infection was identified, lamivudine monotherapy at the HBV daily dose of 100 mg was contraindicated. These patients require treatment for both HIV and HBV with an HIV drug regimen that includes or is supplemented by at least two drugs active against HBV; antiretroviral therapy (ART) often contains two drugs with antiviral activity against HBV (e.g., tenofovir and emtricitabine), but if lamivudine was part of the regimen, the 300-mg daily dose was required (Chap. 202). The safety of lamivudine during pregnancy has not been established; however, the drug is not teratogenic in rodents and has been used safely in pregnant women with HIV infection and with HBV infection. As shown for subsequent nucleoside analogues, administration of lamivudine during the last months of pregnancy to mothers with high-level hepatitis B viremia reduced the likelihood of perinatal transmission of hepatitis B.

ADEFOVIR DIPIVOXIL

At an oral daily dose of 10 mg, the acyclic nucleotide analogue adefovir dipivoxil, the prodrug of adefovir (approved for hepatitis B in 2002), reduces HBV DNA by $\sim 3.5\text{--}4 \log_{10}$ copies/mL, i.e., it is less potent than lamivudine or any of the newer antiviral agents. For a summary of its virologic, serologic, biochemical, and histologic efficacy, as well as its resistance profile, please refer to Table 341-3. Like IFN and lamivudine, adefovir dipivoxil is more likely to achieve an HBeAg response in patients with high baseline ALT; HBeAg responses to it are highly durable and can be relied upon as a treatment stopping point, after a period of consolidation therapy; and biochemical, serologic, and virologic outcomes improve over time with continued therapy.

In HBeAg-negative chronic hepatitis B, as was true for lamivudine, because HBeAg responses—a potential stopping point—cannot be achieved, reactivation is the rule when adefovir therapy is discontinued, and indefinite, long-term therapy is required. Reported attempts to stop adefovir after 5 years were followed by a period of maintained suppression of HBV DNA and ALT; however, most such patients had persistent hepatitis B viremia, and most HBeAg-negative patients were treated indefinitely unless HBsAg loss, albeit very rare, was achieved.

Adefovir contains a flexible acyclic linker instead of the L-nucleoside ring of lamivudine, avoiding steric hindrance by mutated amino acids. In addition, the molecular structure of phosphorylated adefovir is very similar to that of its natural substrate; therefore, mutations to adefovir would also affect binding of the natural substrate, dATP. Thus, resistance to adefovir was much less likely than resistance to lamivudine, and no resistance was encountered in 1 year of clinical trial therapy. In subsequent years, however, adefovir resistance began to emerge (asparagine to threonine at amino acid 236 [N236T] and alanine to valine or threonine at amino acid 181 [A181V/T], primarily), occurring in 2.5% after 2 years but in 29% after 5 years of therapy (reported in HBeAg-negative patients). The primary contribution of adefovir, its effectiveness in lamivudine-resistant, YMDD-mutant HBV, led to its adoption for lamivudine-resistant hepatitis B. When lamivudine resistance occurred, adding adefovir (i.e., maintaining lamivudine to preempt the emergence of adefovir resistance) was superior to switching to adefovir. Almost invariably, patients with adefovir-induced HBV mutations respond to lamivudine (or newer agents, such as entecavir, see below). When, in the past, adefovir had been evaluated as therapy for HIV infection, doses of 60–120 mg were required to suppress HIV, and, at these doses, the drug was nephrotoxic. Even at 30 mg/d, creatinine elevations of 44 μmol/L (0.5 mg/dL) occurred in 10% of patients; however, at the HBV-effective dose of 10 mg, such creatinine elevations were encountered rarely and hardly ever before 6–8 months of therapy. Although renal tubular injury was a rare potential side effect, and although creatinine monitoring was recommended during treatment, the therapeutic index of adefovir dipivoxil was high, and the nephrotoxicity observed in clinical trials at higher doses was reversible. For patients with underlying renal disease, frequency of administration of adefovir had to be reduced, and it could be given only once a week for patients undergoing hemodialysis. Adefovir was very well tolerated, and ALT elevations during and after withdrawal of therapy were similar to those observed and described above in clinical trials of lamivudine. An advantage of adefovir was its relatively favorable resistance profile; however, it was not as potent as the other approved oral agents, it did not suppress HBV DNA as rapidly or as uniformly as the others, it was the least likely of all agents to result in HBeAg seroconversion, and 20–50% of patients failed to suppress HBV DNA by $2 \log_{10}$ ("primary nonresponders"). For these reasons, adefovir, which has been supplanted in both treatment-naïve and lamivudine-resistant patients by the more potent, less resistance-prone tenofovir (see below), is no longer recommended as first-line therapy.

PEGYLATED IFN

After long-acting PEG IFN was shown to be effective in the treatment of hepatitis C (see below), this more convenient IFN preparation was evaluated in the treatment of chronic hepatitis B. Once-a-week PEG IFN is more effective than the more frequently administered, standard IFN, and several large-scale trials of PEG IFN versus oral lamivudine were conducted in patients with chronic hepatitis B.

In HBeAg-reactive chronic hepatitis B, two large-scale studies were done. In one study, PEG IFN- α 2b (100 μg weekly for 32 weeks, then 50 μg weekly for another 20 weeks for a total of 52 weeks) was evaluated against a comparison arm of combination PEG IFN with oral lamivudine in 307 subjects. The other study involved PEG IFN- α 2a (180 μg weekly for 48 weeks) in 814 primarily Asian patients, three-fourths of whom had ALT $\geq 2 \times$ the upper limit of normal, with

comparison arms of lamivudine monotherapy and combination PEG IFN plus lamivudine. At the end of therapy (48–52 weeks) in the PEG IFN monotherapy arms, HBeAg loss occurred in ~30%, HBeAg seroconversion in 22–27%, undetectable HBV DNA (<400 copies/mL by PCR) in 10–25%, and normal ALT in 34–39%, and a mean reduction in HBV DNA of $2 \log_{10}$ copies/mL (PEG IFN- α 2b) to $4.5 \log_{10}$ copies/mL (PEG IFN- α 2a) was seen. Six months after completing PEG IFN monotherapy in these trials, HBeAg losses were present in ~35%, HBeAg seroconversion in ~30%, undetectable HBV DNA in 7–14%, and normal ALT in 32–41%, and the mean reduction in HBV DNA was $2–2.4 \log_{10}$ copies/mL. Although the combination of PEG IFN and lamivudine was superior at the end of therapy in one or more serologic, virologic, or biochemical outcomes, neither the combination arm (in both studies) nor the lamivudine monotherapy arm (in the PEG IFN- α 2a trial) demonstrated any benefit compared to the PEG IFN monotherapy arms 6 months after therapy. Moreover, HBsAg seroconversion occurred in 3–7% of PEG IFN recipients (with or without lamivudine); some of these seroconversions were identified by the end of therapy, but many were identified during the posttreatment follow-up period. The likelihood of HBeAg loss in PEG IFN-treated HBeAg-reactive patients was associated with HBV genotype A > B > C > D (shown for PEG IFN- α 2b but not for PEG IFN- α 2a). PEG IFN- α 2a was approved in the United States for hepatitis B in 2005; PEG IFN- α 2b, which is not approved for hepatitis B in the United States, is used in other countries.

Based on these results, some authorities concluded that PEG IFN monotherapy should be the first-line therapy of choice in HBeAg-reactive chronic hepatitis B; however, this conclusion has been challenged. Although a finite, 1-year course of PEG IFN results in a higher rate of sustained response (6 months after treatment) than is achieved with oral nucleoside/nucleotide analogue therapy, the comparison is confounded by the fact that oral agents are not discontinued at the end of 1 year. Instead, taken orally and free of side effects, therapy with oral agents is extended indefinitely or until after the occurrence of an HBeAg response. The rate of HBeAg responses after 2 years of oral-agent nucleoside analogue therapy is at least as high as, if not higher than, that achieved with PEG IFN after 1 year; favoring oral agents is the absence of injections, difficult-to-tolerate side effects, and laboratory monitoring as well as lower direct and indirect medical care costs and inconvenience. The association of HBsAg responses with PEG IFN therapy occurs in such a small proportion of patients that subjecting everyone to PEG IFN for the marginal gain of HBsAg responses during or immediately after therapy in such a very small minority is questionable. Moreover, HBsAg responses occur in a comparable proportion of patients treated with early-generation nucleoside/nucleotide analogues in the years *after* therapy, and, with the newer, more potent nucleoside analogues, the frequency of HBsAg loss during the first year of therapy equals that of PEG IFN and is exceeded during year 2 and beyond (see below). Of course, resistance is not an issue during PEG IFN therapy, but the risk of resistance is much lower with new agents (≤1% up to 3–8 years in previously treatment-naïve, entecavir-treated patients and 0% in tenofovir-treated patients; see below). Finally, the level of HBV DNA inhibition that can be achieved with the newer agents, and even with lamivudine, exceeds that achieved with PEG IFN, in some cases by several orders of magnitude.

In HBeAg-negative chronic hepatitis B, a trial of PEG IFN- α 2a (180 μg weekly for 48 weeks vs comparison arms of lamivudine monotherapy and of combination therapy) in 564 patients showed that PEG IFN monotherapy resulted at the end of therapy in suppression of HBV DNA by a mean of $4.1 \log_{10}$ copies/mL, undetectable HBV DNA (<400 copies/mL by PCR) in 63%, normal ALT in 38%, and loss of HBsAg in 4%. Although lamivudine monotherapy and combination lamivudine–PEG IFN therapy were both superior to PEG IFN at the end of therapy, no advantage of lamivudine monotherapy or combination therapy was apparent over PEG IFN monotherapy 6 months after therapy—suppression of HBV DNA by a mean of $2.3 \log_{10}$ copies/mL, undetectable HBV DNA in 19%,

and normal ALT in 59%. In patients involved in this trial followed for up to 5 years, among the two-thirds followed who had been treated initially with PEG IFN, 17% maintained HBV DNA suppression to <400 copies/mL, but ALT remained normal in only 22%; HBsAg loss increased gradually to 12%. Among the half followed who had been treated initially with lamivudine monotherapy, HBV DNA remained <400 copies/mL in 7% and ALT normal in 16%; by year 5, 3.5% had lost HBsAg. As was the case for standard IFN therapy in HBeAg-negative patients, only a small proportion maintained responsiveness after completion of PEG IFN therapy, raising questions about the relative value of a finite period of PEG IFN versus a longer course with a potent, low-resistance oral nucleoside analogue in these patients. Moreover, the value of PEG IFN for HBeAg-negative chronic hepatitis B has not been confirmed. In the only other controlled clinical trial of PEG IFN for HBeAg-negative chronic hepatitis B, the hepatitis C regimen of PEG IFN plus ribavirin was compared to PEG IFN monotherapy. In this trial, HBV DNA suppression (<400 copies/mL) occurred in only 7.5% of the two groups combined, and no study subject lost HBsAg.

In patients treated with PEG IFN, HBeAg and HBsAg responses have been associated with *IL28B* (now renamed IFN lambda-3, *IFNL3*) genotype CC, the favorable genotype identified in trials of PEG IFN for chronic hepatitis C. Also, reductions in quantitative HBsAg levels have been shown to correlate with and to be predictive of responsiveness to PEG IFN in chronic hepatitis B. If HBsAg levels fail to fall within the first 12–24 weeks or to reach <20,000 IU/mL by week 24, PEG IFN therapy is unlikely to be effective and should be discontinued. (Similar observations of HBsAg levels in oral-agent-treated patients are of interest but of limited clinical relevance, given the very high likelihood of virologic responses during such therapy.) While PEG IFN remains one of the recommended first-line agents for hepatitis B, subsequent-generation, injection-free, very-well-tolerated, high-barrier-to-resistance, oral agents are used much more widely.

ENTECAVIR

Entecavir, an oral cyclopentyl guanosine analogue polymerase inhibitor (approved in 2005), appears to be the most potent of the HBV antivirals and is just as well tolerated as lamivudine. In a 709-subject clinical trial among HBeAg-reactive patients, oral entecavir, 0.5 mg daily, was compared to lamivudine, 100 mg daily. At 48 weeks, entecavir was superior to lamivudine in suppression of HBV DNA (mean 6.9 vs 5.5 log₁₀ copies/mL), percentage with undetectable HBV DNA (<300 copies/mL by PCR; 67 vs 36%), histologic improvement (2-point improvement in necroinflammatory HAI score; 72 vs 62%), and normal ALT (68 vs 60%). The two treatments were indistinguishable in percentage with HBeAg loss (22 vs 20%) and seroconversion (21 vs 18%). Among patients treated with entecavir for 96 weeks, HBV DNA was undetectable cumulatively in 80% (vs 39% for lamivudine), and HBeAg seroconversions had occurred in 31% (vs 26% for lamivudine). After 3–6 years of entecavir, HBeAg seroconversions were observed in 39–44% and HBsAg loss in 5–6%. Similarly, in a 638-subject clinical trial among HBeAg-negative patients, at week 48, oral entecavir, 0.5 mg daily, was superior to lamivudine, 100 mg daily, in suppression of HBV DNA (mean 5.0 vs 4.5 log₁₀ copies/mL) and in percentage with undetectable HBV DNA (90 vs 72%), histologic improvement (70 vs 61%), and normal ALT (78 vs 71%). No resistance mutations were encountered in previously treatment-naïve, entecavir-treated patients during 96 weeks of therapy, and in a cohort of subjects treated for up to 6 years, resistance emerged in only 1.2%. Entecavir-induced HBeAg seroconversions are as durable as those achieved with other antivirals. Its high barrier to resistance coupled with its high potency renders entecavir a first-line drug for patients with chronic hepatitis B.

Entecavir is also effective against lamivudine-resistant HBV infection. In a trial of 286 lamivudine-resistant patients, entecavir, at a higher daily dose of 1 mg, was superior to lamivudine, as measured at week 48, in achieving suppression of HBV DNA (mean

5.1 vs 0.48 log₁₀ copies/mL), undetectable HBV DNA (72 vs 19%), normal ALT (61 vs 15%), HBeAg loss (10 vs 3%), and HBeAg seroconversion (8 vs 3%). In this population of lamivudine-experienced patients, however, entecavir resistance emerged in 7% at 48 weeks. Although entecavir resistance requires both a YMDD mutation and a second mutation at one of several other sites (e.g., T184A, S202G/I, or M250V), resistance to entecavir in lamivudine-resistant chronic hepatitis B was reported to increase progressively to 43% at 4 years and 57% at 6 years; therefore, entecavir is not as attractive a choice (and is not recommended, despite its approval for this indication) as adefovir was or as tenofovir is for patients with lamivudine-resistant hepatitis B.

In clinical trials, entecavir had an excellent safety profile. In addition, on-treatment and posttreatment ALT flares are relatively uncommon and relatively mild in entecavir-treated patients. Doses should be reduced for patients with reduced creatinine clearance. Entecavir does have low-level antiviral activity against HIV and cannot be used as monotherapy to treat HBV infection in HIV/HBV co-infected persons.

TELBIVUDINE

Telbivudine, a cytosine analogue (approved in 2006), is similar in efficacy to entecavir but slightly less potent in suppressing HBV DNA (a slightly less profound median 6.4 log₁₀ reduction in HBeAg-reactive disease and a similar 5.2 log₁₀ reduction in HBeAg-negative disease). In its registration trial, telbivudine at an oral daily dose of 600 mg suppressed HBV DNA to <300 copies/mL in 60% of HBeAg-positive and 88% of HBeAg-negative patients, reduced ALT to normal in 77% of HBeAg-positive and 74% of HBeAg-negative patients, and improved histology in 65% of HBeAg-positive and 67% of HBeAg-negative patients. Although resistance to telbivudine (M204I, not M204V, mutations) was less frequent than resistance to lamivudine at the end of 1 year, resistance mutations after 2 years of treatment occurred in up to 22%. Generally well tolerated, telbivudine was associated with a low frequency of asymptomatic creatine kinase elevations and with a very low frequency of peripheral neuropathy; frequency of administration had to be reduced for patients with impaired creatinine clearance. Its excellent potency notwithstanding, the inferior resistance and safety profile of telbivudine limited its appeal; telbivudine is neither recommended as first-line therapy nor widely used.

TENOFOVIR

Tenofovir disoproxil fumarate (TDF), an acyclic nucleotide analogue and potent antiretroviral agent used to treat HIV infection (approved for hepatitis B in 2008), is similar to adefovir but more potent in suppressing HBV DNA and inducing HBeAg responses; it is highly active against both wild-type and lamivudine-resistant HBV and active in patients whose response to adefovir is slow and/or limited. At an oral once-daily dose of 300 mg for 48 weeks, tenofovir suppressed HBV DNA by 6.2 log₁₀ (to undetectable levels [<400 copies/mL] in 76%) in HBeAg-positive patients and by 4.6 log₁₀ (to undetectable levels in 93%) in HBeAg-negative patients; reduced ALT to normal in 68% of HBeAg-positive and 76% of HBeAg-negative patients; and improved histology in 74% of HBeAg-positive and 72% of HBeAg-negative patients. In HBeAg-positive patients, HBeAg seroconversions occurred in 21% by the end of year 1, 27% by year 2, 34% by year 3, and 40% by year 5 of tenofovir treatment; HBsAg loss occurred in 3% by the end of year 1, 6% at year 2, and 8% by year 5. After 5 years of tenofovir therapy, 87% of patients experienced histologic improvement, including reduction in fibrosis score (51%) and regression of cirrhosis (71%). The 5-year safety (negligible renal toxicity, in 1%, and mild reduction in bone density, in ~0.5%) and resistance profiles (none recorded through 8 years) of tenofovir are very favorable as well; therefore, tenofovir has supplanted adefovir both as first-line therapy for chronic hepatitis B and as rescue therapy for lamivudine-resistant chronic hepatitis B. Studies of tenofovir and entecavir reviewed in 2015 showed no difference in long-term risks of renal and bone

toxicity; however, among patients treated with tenofovir, instances of acute renal failure and of low blood phosphate levels have been reported. Thus, in patients receiving tenofovir, monitoring bone density is not recommended, but periodic (at least annual) monitoring for renal injury is recommended (serum creatinine and phosphate, urine glucose and protein). Frequency of tenofovir administration should be reduced for patients with impaired creatinine clearance.

Tenofovir alafenamide (TAF), a second-generation tenofovir approved in 2016, is a prodrug of tenofovir that requires activation to tenofovir in hepatocytes. This targeted delivery to hepatocytes allows a lower dose to suffice and reduces systemic exposure by 90%, thereby minimizing TDF-associated proximal tubular renal injury, its associated phosphate wasting, and the potential consequent loss of bone mineral density. The dose of TAF is 25 mg, which is equivalent in antiviral potency to 300 mg of TDF; both formulations have the same high barrier to resistance, and clinical resistance has not been encountered. Randomized, controlled, double-blind, phase 3 noninferiority trials, one in HBeAg-positive patients and the other in HBeAg-negative patients, provided the safety and efficacy data to support TAF approval.

In 873 *HBeAg-positive* patients treated for 48 weeks, TAF versus TDF achieved (1) HBV DNA reductions to <29 IU/mL in 64% versus 67%; (2) ALT normalization in 72% versus 67% (an unexplained TAF biochemical advantage confirmed in other trials); (3) HBeAg loss in 14% versus 12%; (4) HBeAg seroconversion in 10% versus 8%; and (5) a negligible loss of HBsAg in 1% versus 0.3%. Compared to TDF, TAF was associated with reduced impairment of renal function (median reduction in estimated glomerular filtration rate of -0.6 mL/min for TAF vs -5.4 mL/min for TDF) and of bone density (in hip measurements, mean reduction of -0.10% for TAF vs -1.72% for TDF; adjusted difference, 1.62%).

In the parallel trial among 426 *HBeAg-negative* patients treated for 48 weeks, reductions in HBV DNA to <29 IU/mL occurred in 94% versus 93% of individuals treated with TAF versus TDF, respectively; normalization of ALT occurred in 83% versus 75%, but no HBsAg loss occurred in either group. Similar TAF advantages in maintaining renal function and bone density were reported: reduction in median estimated glomerular filtration rate (-1.8 mL/min for TAF vs -4.8 mL/min for TDF) and in median bone density (in hip measurements, mean reduction of -0.29% for TAF vs -2.16% for TDF; adjusted percentage difference, 1.87%).

At week 96, TAF and TDF HBV DNA and ALT reductions (including the TAF advantage observed at 48 weeks) were maintained. In the original TDF group, when TDF was switched to TAF after week 96, all differences observed during the first 96 weeks (in normalization of ALT and reductions in renal function and bone density) had resolved at week 120. Resistance did not emerge to either TAF or TDF throughout the trial.

Based on these trial outcomes, TAF joined the list of recommended first-line antiviral agents for chronic hepatitis B. This drug is recommended over TDF by the American Association for the Study of Liver Diseases (AASLD) and the European Association for the Study of the Liver (EASL) for patients with reduced renal function (creatinine clearance <50 mL/min), reduced bone density, and risk factors for renal injury (including, according to EASL guidelines, decompensated cirrhosis, creatinine clearance <60 mL/min, poorly controlled hypertension or diabetes, proteinuria, active glomerulonephritis, concomitant nephrotoxic medications, or solid-organ transplantation); the EASL recommendation extends to persons >60 years, who are at increased risk of TDF nephrotoxicity. In patients with creatinine clearances <15 mL/min, neither TDF nor TAF is recommended.

A comparison of antiviral therapies for chronic hepatitis B appears in Table 341-3; their relative potencies in suppressing HBV DNA are shown in Fig. 341-1.

COMBINATION THERAPY

Although the combination of lamivudine and PEG IFN suppresses HBV DNA more profoundly during therapy than does

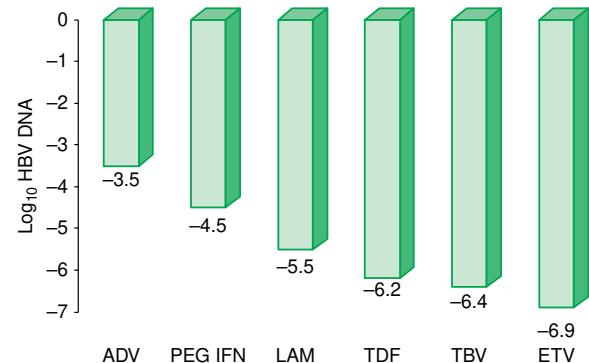


FIGURE 341-1 Relative potency of antiviral drugs for hepatitis B, as reflected by median \log_{10} hepatitis B virus (HBV) DNA reduction in HBeAg-positive chronic hepatitis B. These data are from individual reports of large, randomized controlled registration trials that were the basis for approval of the drugs. In most instances, these data do not represent direct comparisons among the drugs, because study populations were different, baseline patient variables were not always uniform, and the sensitivity and dynamic range of the HBV DNA assays used in the trials varied. ADV, adefovir dipivoxil; ETV, entecavir; LAM, lamivudine; PEG IFN, pegylated interferon α 2a; TBV, telbivudine; TDF, tenofovir disoproxil fumarate. Because of potency and a high barrier to resistance, ETV and tenofovir (either TDF or the second-generation tenofovir alafenamide) are recommended as first-line therapy. While PEG IFN remains a first-line agent, the oral agents developed earlier, LAM, ADV, and TBV, are no longer preferred agents.

monotherapy with either drug alone (and is much less likely to be associated with lamivudine resistance), this combination used for a year is no better than a year of PEG IFN in achieving sustained responses. To date, combinations of oral nucleoside/nucleotide agents have not achieved an enhancement in virologic, serologic, or biochemical efficacy over that achieved by the more potent of the combined drugs given individually. In a 2-year trial of combination entecavir and tenofovir versus entecavir monotherapy, for a small subgroup of patients with very high HBV DNA levels ($\geq 10^8$ IU/mL), a reduction in HBV DNA to <50 IU/mL was higher in the combination group (79 vs 62%); however, no differences in HBeAg responses or any other endpoint were observed between the combination-therapy and monotherapy groups, even in the high-HBV DNA subgroup. For resistance to lamivudine or adefovir, adding a second, non-cross-resistant agent was the chosen approach. Whereas, initially, in clinical studies of adefovir as rescue therapy for lamivudine resistance, adding adefovir to lamivudine (combination therapy) was considered a better strategy than replacing lamivudine with adefovir monotherapy (to minimize ALT flares and to avoid adefovir resistance), according to current treatment recommendations of the AASLD and the EASL, switching from the resistant drug to the new drug is preferred. Because the current generation of antivirals is so potent and has such a high barrier to resistance, monotherapy with the rescue drug (e.g., tenofovir for lamivudine resistance) is as effective (as demonstrated in observational reports for up to 5 years) in maintaining viral suppression without the emergence of resistance as combination therapy with the resistant drug and the rescue drug. Generally, in patients treated with entecavir and tenofovir preparations, antiviral drug resistance is no longer encountered. For currently rare patients who already have acquired multidrug resistance (to both nucleoside analogues [lamivudine, entecavir, telbivudine] and nucleotide analogues [adefovir, tenofovir]), treatment with a combination of entecavir and tenofovir has been shown to be highly effective in suppressing HBV DNA and overcoming drug resistance.

NOVEL ANTIVIRALS AND STRATEGIES

In addition to the eight approved antiviral drugs for hepatitis B, emtricitabine, a fluorinated cytosine analogue very similar to lamivudine in structure, efficacy, and resistance profile, offers no advantage over lamivudine. A combination of emtricitabine and tenofovir is approved for the treatment of HIV infection and is

an appealing combination therapy for hepatitis B, especially for lamivudine-resistant disease; however, neither emtricitabine nor the combination is approved for hepatitis B. Several initially promising antiviral agents have been abandoned because of toxicity (e.g., clevudine, which was linked to myopathy during its clinical development). The current generation of oral antivirals have been very successful in the management of chronic hepatitis B; however, most patients require long-duration, usually indefinite, therapy. Ideally, an approach to achieving “cure” (eradication of HBV infection) with finite-duration therapy would be welcome. Currently, innovative approaches being investigated focus on viral-targeting strategies or immunomodulatory strategies. The direct viral approaches include viral entry inhibitors, nucleocapsid assembly inhibitors, HBV secretion (HBsAg release) inhibitors, covalently closed circular (ccc) DNA silencing/inhibition/cleavage, RNA interference, HBx inhibitors, and CRISPR/Cas9 gene editing. Immunomodulators being studied have included Toll receptor agonists, T-cell vaccines, programmed cell death 1 (PD-1) blockade, reconstitution of innate and adaptive immune responses, and HBV mRNA recognition and activation of innate immune signaling by retinoic acid–inducible gene-I (RIG-I). While data supporting several of these unconventional approaches have begun to appear, some have been abandoned for lack of efficacy or for toxicity (e.g., RIG-I and some of the capsid inhibitors). Even after almost a decade of early clinical trials, none has been shown to “cure” hepatitis B, and none is likely to be competitive, unless it can be shown to go beyond current antivirals in achieving recovery (HBsAg seroconversion) from HBV infection.

TREATMENT RECOMMENDATIONS

Several learned societies and groups of expert physicians have issued treatment recommendations for patients with chronic hepatitis B; the most authoritative and updated are those of the AASLD and the

EASL. Although the recommendations differ slightly, a consensus has emerged on most of the important points (Table 341-4). No treatment is recommended or available for inactive “nonreplicative” hepatitis B carriers (undetectable HBeAg with normal ALT and HBV DNA $\leq 10^3$ IU/mL documented serially over time). In patients with detectable HBeAg and HBV DNA levels $>2 \times 10^4$ IU/mL, treatment is recommended by the AASLD for those with ALT levels $>2 \times$ the upper limit of normal. (The EASL recommends treatment in HBeAg-positive patients for HBV DNA levels $>2 \times 10^3$ IU/mL and ALT above the upper limit of normal.) For HBeAg-positive patients with ALT $\leq 2 \times$ the upper limit of normal, in whom sustained responses are not likely and who would require multiyear therapy, antiviral therapy is not recommended currently. This pattern is common during the early decades of life among Asian patients infected at birth; even in this group, therapy would be considered for those >40 years of age, patients with extrahepatic manifestations of HBV infection, patients with a family history of cirrhosis or HCC, if the liver biopsy or noninvasive testing shows moderate to severe necroinflammatory activity or fibrosis, or if the patient has a history of previous treatment. In this group, when, eventually, ALT becomes elevated later in life, antiviral therapy should be instituted. For patients with HBeAg-negative chronic hepatitis B, ALT $>2 \times$ the upper limit of normal (above the upper limit of normal according to EASL), and HBV DNA $>2 \times 10^3$ IU/mL, antiviral therapy is recommended. If HBV DNA is $>2 \times 10^3$ IU/mL and ALT is 1 to $>2 \times$ the upper limit of normal, the same considerations apply as for HBeAg-positive patients with borderline ALT levels—for those >40 years of age, patients with extrahepatic manifestations of HBV infection, patients with a family history of cirrhosis or HCC, if the liver biopsy or noninvasive testing shows moderate to severe necroinflammatory activity or fibrosis, or if the patient has a history of previous treatment (treatment in this subset would be

TABLE 341-4 Recommendations for Treatment of Chronic Hepatitis B^a

HBeAg STATUS	CLINICAL	HBV DNA (IU/mL)	ALT	RECOMMENDATION
HBeAg-reactive	Chronic hepatitis	$>2 \times 10^4$	$\leq 2 \times \text{ULN}^{c,d}$	No treatment; monitor, except in patients >40 , with family history of cirrhosis or hepatocellular carcinoma, with extrahepatic manifestations, with a history of previous treatment, and/or with liver biopsy (or noninvasive fibrosis determination) evidence for moderate to severe inflammation or fibrosis
		$>2 \times 10^4$	$>2 \times \text{ULN}^d$	Treat ^b
		$>2 \times 10^3$	$< \text{or} > \text{ULN}$	Treat ^b with oral agents, not PEG IFN
	Cirrhosis decompensated	$<2 \times 10^3$	$>\text{ULN}$	Treatment suggested ^d
HBeAg-negative	Chronic hepatitis	Detectable	$< \text{or} > \text{ULN}$	Treat ^b with oral agents ^e , not PEG IFN; refer for liver transplantation
		Undetectable	$< \text{or} > \text{ULN}$	Observe; refer for liver transplantation
	Cirrhosis compensated	$\leq 2 \times 10^3$	$\leq \text{ULN}$	Inactive carrier; treatment not necessary
		$>2 \times 10^3$	$1 \text{ to } > 2 \times \text{ULN}^d$	No treatment; monitor, except in patients >40 , with family history of cirrhosis or hepatocellular carcinoma, with extrahepatic manifestations, with a history of previous treatment, and/or with liver biopsy (or noninvasive fibrosis determination) evidence for moderate to severe inflammation or fibrosis
		$>2 \times 10^3$	$>2 \times \text{ULN}^d$	Treat ^{b,i}
	Cirrhosis decompensated	$>2 \times 10^3$	$< \text{or} > \text{ULN}$	Treat ^b with oral agents, not PEG IFN
		$<2 \times 10^3$	$>\text{ULN}$	Treatment suggested ^d
		Detectable	$< \text{or} > \text{ULN}$	Treat ^b with oral agents ^e , not PEG IFN; refer for liver transplantation
		Undetectable	$< \text{or} > \text{ULN}$	Observe; refer for liver transplantation

^aBased on practice guidelines of the American Association for the Study of Liver Diseases (AASLD). Except as indicated in footnotes, these guidelines are similar to those issued by the European Association for the Study of the Liver (EASL). ^bLiver disease tends to be mild or inactive clinically; most such patients do not undergo liver biopsy.

^cThis pattern is common during early decades of life in Asian patients infected at birth. ^dAccording to the EASL guidelines, treat if HBV DNA is $>2 \times 10^3$ IU/mL and ALT $>\text{ULN}$.

^eOne of the potent oral drugs with a high barrier to resistance (entecavir or tenofovir) or PEG IFN can be used as first-line therapy (see text). These oral agents, but not PEG IFN, should be used for interferon-refractory/intolerant and immunocompromised patients. PEG IFN is administered weekly by subcutaneous injection for a year; the oral agents are administered daily for at least a year and continued indefinitely or until at least 6 months after HBeAg seroconversion. ^fAccording to EASL guidelines, patients with compensated cirrhosis and detectable HBV DNA at any level, even with normal ALT, are candidates for therapy. Most authorities would treat indefinitely, even in HBeAg-positive disease after HBeAg seroconversion. ^gBecause the emergence of resistance can lead to loss of antiviral benefit and further deterioration in decompensated cirrhosis, a low-resistance regimen is recommended—entecavir or tenofovir monotherapy or combination therapy with the more resistance-prone lamivudine (or telbivudine) plus adefovir. Therapy should be instituted urgently. ^hBecause HBeAg seroconversion is not an option, the goal of therapy is to suppress HBV DNA and maintain a normal ALT. PEG IFN is administered by subcutaneous injection weekly for a year; caution is warranted in relying on a 6-month posttreatment interval to define a sustained response, because the majority of such responses are lost thereafter. Oral agents, entecavir or tenofovir, are administered daily, usually indefinitely or until, as very rarely occurs, virologic and biochemical responses are accompanied by HBsAg seroconversion. ⁱFor older patients and those with advanced fibrosis, consider lowering the HBV DNA threshold to $>2 \times 10^3$ IU/mL.

Abbreviations: ALT, alanine aminotransferase; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HBV, hepatitis B virus; PEG IFN, pegylated interferon; ULN, upper limit of normal.

recommended according to EASL guidelines, because ALT is elevated). Per current AASLD recommendations, antiviral treatment with oral agents can be stopped after HBeAg seroconversion in noncirrhotics, and the suggested period of consolidation therapy is 12 months with close monitoring for recurrent viremia (monthly \times 6, then every 3 months for the rest of a year) after cessation of therapy. For patients with HBeAg-negative chronic hepatitis, the current recommendation with oral agents is for indefinite therapy; stopping therapy in this group can be considered after HBsAg loss.

The potential for stopping antiviral therapy in noncirrhotic HBeAg-negative patients after protracted (≥ 5 years) antiviral therapy has been the subject of several studies. After such prolonged courses of entecavir or tenofovir, in one study (DARING-B), 18-month virologic relapse rates (HBV DNA >2000 IU/mL) in 57 patients were high (in 72%), but only 26% met study criteria for resumption of therapy (ALT $>10 \times$ upper limit of normal, ALT $>5 \times$ upper limit of normal with bilirubin >2 mg/mL, ALT $>3 \times$ upper limit of normal with HBV DNA $>10^5$ IU/mL, or ALT $>2 \times$ upper limit of normal and HBV DNA $>2 \times 10^3$ IU/mL on three sequential visits). Moreover, 25% underwent HBsAg loss. In a similar study (FINITE), virologic relapse rates were high, but 62% did not meet criteria for retreatment, and 19% lost HBsAg. In contrast, in a study among Asian patients, only ~30% had sustained responses for which resumption of therapy was not introduced, and HBsAg responses were negligible. In other reports, including on patients with 8 years of TDF treatment prior to stopping, only 35–60% had sustained treatment-free outcomes, and only 5–13% lost HBsAg. In the only randomized, controlled trial of stopping therapy versus continuing therapy in HBeAg-negative patients after prolonged antiviral therapy (Toronto STOP study), only 33% had sustained responses after cessation of therapy, and HBsAg loss occurred with equal, small frequencies in both the stop-treatment group (4%) and the continue-treatment group (5%). Generally, then, although HBsAg loss can be achieved in a small fraction and although a subgroup may not require reintroduction of therapy in the short run, enthusiasm for this approach is limited, and for most HBeAg-negative patients, recommendations support indefinite treatment, unless they experience HBsAg loss.

For patients with compensated cirrhosis, because antiviral therapy has been shown to retard clinical progression, treatment is recommended regardless of HBeAg status and ALT as long as HBV DNA is detectable at $>2 \times 10^3$ IU/mL (detectable at any level according to the EASL); therapy is suggested, however, even for those with HBV DNA $<2 \times 10^3$ IU/mL, regardless of ALT level. For patients with decompensated cirrhosis, treatment is recommended regardless of serologic and biochemical status, as long as HBV DNA is detectable. Patients with decompensated cirrhosis should be evaluated as candidates for liver transplantation. Cirrhotics should be treated indefinitely (see considerations for stopping antiviral therapy in noncirrhotics, above).

Among the eight available drugs for hepatitis B, PEG IFN has supplanted standard IFN, entecavir has supplanted lamivudine, and tenofovir has supplanted adefovir. PEG IFN, entecavir, or tenofovir (TDF or TAF) is recommended as first-line therapy (Table 341-3). PEG IFN requires finite-duration therapy, achieves the highest rate of HBeAg responses after a year of therapy, and does not support viral mutations, but it requires subcutaneous injections and is associated with inconvenience, more intensive clinical and laboratory monitoring, and intolerance. Oral nucleoside analogues require long-term therapy in most patients, and when used alone, lamivudine and telbivudine foster the emergence of viral mutations, adefovir somewhat less so, and entecavir (except in lamivudine-experienced patients) and tenofovir rarely at all. Oral agents do not require injections or cumbersome laboratory monitoring, are very well tolerated, lead to improved histology in 50–90% of patients, suppress HBV DNA more profoundly than PEG IFN, and are effective even in patients who fail to respond to IFN-based therapy. Although oral agents are less likely to result in HBeAg responses during the first year of therapy, as compared to PEG IFN, treatment

TABLE 341-5 Pegylated Interferon Versus Oral Nucleoside Analogues for the Treatment of Chronic Hepatitis B

	PEG IFN	NUCLEOSIDE ANALOGUES
Administration	Weekly injection	Daily, orally
Tolerability	Poorly tolerated, intensive monitoring	Well tolerated, limited monitoring
Duration of therapy	Finite 48 weeks	≥ 1 year, indefinite in most patients
Maximum mean HBV DNA suppression	$4.5 \log_{10}$	$6.9 \log_{10}$
Effective in high-level HBV DNA ($\geq 10^6$ IU/mL)	No	Yes
HBeAg seroconversion	~30%	~20%
During 1 year of therapy	Not applicable	30% (year 2) to up to 50% (year 5)
During >1 year of therapy		
HBeAg-negative posttreatment HBV DNA suppression	17% at 5 years	7% at 4 years (lamivudine)
HBsAg loss	3–4%	0–3%
During 1 year of therapy	Not applicable	3–8% at 5 years of therapy
During >1 year of therapy	12% at 5 years	3.5% at 5 years
After 1 year of therapy—HBeAg-negative		
Antiviral resistance	None	Lamivudine: ~30% year 1, ~70% year 5 Adefovir: 0% year 1, ~30% year 5 Telbivudine: up to 4% year 1, 22% year 2 Entecavir: $\leq 1.2\%$ through year 6 Tenofovir: 0% through year 8
Use in cirrhosis, transplantation, immunosuppressed	No	Yes
Cost, 1 year of therapy	++++	+ to ++

Abbreviations: HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; PEG IFN, pegylated interferon.

with oral agents is usually extended beyond the first year and, by the end of the second year (of the current generation of potent agents entecavir and tenofovir), yields HBeAg responses (and even HBsAg responses) comparable in frequency to those achieved after 1 year of PEG IFN (and without the associated side effects) (Table 341-5). In a 2016 systematic review of 1716 patients involved in 25 clinical trials, responses after oral-agent therapy were found to be durable. Among patients with HBeAg-reactive chronic hepatitis B, the pooled rates of durable HBeAg seroconversions maintained after cessation of nucleoside/nucleotide analogue therapy (including all the oral agents) were 92% and 88% at posttreatment months 12 and 24, respectively, unaffected by the duration of post-HBeAg-response consolidation therapy (>6 months in all studies evaluated); the pooled rate of durable biochemical remission after therapy in this population was 76%. Even for HBeAg-negative chronic hepatitis B, for which most authorities recommend indefinite therapy, pooled rates of virologic remissions maintained after cessation of oral-agent therapy were 44%, 31%, and 30% at posttreatment months 12, 24, and 36, and the pooled rate of durable biochemical remission in this population was 57%.

Although adefovir and tenofovir (TDF) are safe, renal monitoring (e.g., serum creatinine and phosphate, urine glucose and protein) is recommended (not for TAF). Substantial experience with lamivudine during pregnancy (see above) has identified no teratogenicity; although widely used during pregnancy, lamivudine remains classified as pregnancy category C. Although IFNs do not

appear to cause congenital anomalies, these have antiproliferative properties and should be avoided during pregnancy. Adefovir during pregnancy has not been associated with birth defects; however, the risk of spontaneous abortion may be increased, and adefovir is categorized as pregnancy category C. Data on the safety of entecavir during pregnancy have not been published (pregnancy category C). Sufficient data in animals and limited data in humans suggest that telbivudine and tenofovir (both pregnancy category B) can be used safely during pregnancy; however, telbivudine is not an acceptable first-line drug. In general, then, except for lamivudine and tenofovir, and until additional data become available, the other antivirals for hepatitis B should be avoided or used with extreme caution during pregnancy. Tenofovir is the current drug of choice in pregnancy.

For children aged 2 to <18 years old with HBeAg-reactive hepatitis B (most children will be HBeAg-reactive; no studies have been done in children with HBeAg-negative chronic hepatitis B), treatment is recommended if HBV DNA is detectable and ALT levels are elevated, but not if ALT levels are normal. Each of the available drugs, except telbivudine, is approved for different childhood age groups (standard IFN α 2b age \geq 1 year; PEG IFN α 2a age \geq 5 years [approved for hepatitis C, not B, but can be used in hepatitis B]; lamivudine and entecavir age \geq 2 years; adefovir and tenofovir age \geq 2 years). Package inserts should be consulted for childhood doses.

As noted above, some physicians prefer to begin with PEG IFN, while other physicians and patients prefer oral agents as first-line therapy. For patients with decompensated cirrhosis, the emergence of resistance can result in further deterioration and loss of antiviral effectiveness. Therefore, in this patient subset, therapy with a very favorable resistance profile (e.g., entecavir or tenofovir) should be used. PEG IFN should not be used in patients with compensated or decompensated cirrhosis.

Several observational studies have suggested that TDF is superior to entecavir in reducing the risk of HCC. Such studies, however, sophisticated statistical analyses notwithstanding, are subject to confounding influences that could favor TDF; in addition, while several studies confirm a differential effect of TDF on long-term HCC risk, many others do not. Therefore, currently, the preponderance of data is insufficient to support this benefit of TDF over entecavir.

For patients with end-stage chronic hepatitis B who undergo liver transplantation, reinfection of the new liver is almost universal in the absence of antiviral therapy. The majority of patients become high-level viremic carriers with minimal liver injury. Before the availability of antiviral therapy, an unpredictable proportion experienced severe hepatitis B-related liver injury, sometimes a fulminant-like hepatitis and sometimes a rapid recapitulation of the original severe chronic hepatitis B (*Chap. 339*). Currently, however, prevention of recurrent hepatitis B after liver transplantation has been achieved definitively by combining short-term (5–7 days) intravenous hepatitis B immune globulin (HBIG) with lifelong low-resistance oral entecavir or TDF or TAF (*Chap. 345*); in some patients, especially those with a low risk for recurrence, the newer, more potent, and less resistance-prone oral agents may be used instead of HBIG for posttransplantation therapy. For patients at high risk for recurrence and progressive disease (e.g., patients with HDV-HBV or HIV-HBV co-infection as well as for nonadherent patients, lifelong combination HBIG-oral agent therapy should be considered. For patients receiving livers from anti-HBc-positive donors, lifelong oral-agent therapy is recommended (without HBIG).

Patients with HBV-HIV co-infection can have progressive HBV-associated liver disease and, occasionally, a severe exacerbation of hepatitis B resulting from immunologic reconstitution following ART. Lamivudine should never be used as monotherapy in patients with HBV-HIV infection because HIV resistance emerges rapidly to both viruses. Adefovir was used successfully in the past to treat chronic hepatitis B in HBV-HIV co-infected patients but is no longer considered a first-line agent for HBV. Entecavir has low-level

activity against HIV and can result in selection of HIV resistance; therefore, it is not preferable in HBV-HIV co-infection. Tenofovir and the combination of tenofovir and emtricitabine in one pill are approved therapies for HIV and represent excellent choices for treating HBV infection in HBV-HIV co-infected patients. Generally, even for HBV-HIV co-infected patients who do not yet meet treatment criteria for HIV infection, treating for both HBV and HIV is recommended. In HIV-HBV co-infection, TAF is preferable to TDF because of its better safety profile.

Patients with chronic hepatitis B who undergo cytotoxic chemotherapy for treatment of malignancies as well as patients treated with immunosuppressive, anticytokine, or anti-tumor necrosis factor (TNF) therapies (the risk varies, from highest [e.g., B cell-depleting agents, anthracycline derivatives, moderate-/high-dose corticosteroids for \geq 4 weeks] to moderate [e.g., TNF- α inhibitors, cytokine or integrin inhibitors, tyrosine kinase inhibitors, low-dose corticosteroids for \geq 4 weeks] to lowest [e.g., immunosuppressive agents like methotrexate and azathioprine, intraarticular corticosteroids, any dose of corticosteroids for \leq 1 week]) experience enhanced HBV replication and viral expression on hepatocyte membranes during chemotherapy coupled with suppression of cellular immunity. When chemotherapy is withdrawn, such patients are at risk for reactivation of hepatitis B, often severe and occasionally fatal. Such rebound reactivation represents restoration of cytolytic T-cell function against a target organ enriched in HBV expression. Preemptive treatment with the first of the oral HBV antivirals, lamivudine, prior to the initiation of chemotherapy was shown to reduce the risk of such reactivation substantially; treating *after* reactivation has occurred is less effective. The newer, more potent oral antiviral agents, entecavir and tenofovir, which are even more effective in preventing hepatitis B reactivation and with a lower risk of antiviral drug resistance, are preferred. The optimal duration of antiviral therapy after completion of chemotherapy is not known, but a suggested approach is 6 months (12 months for B cell-depleting agents) for inactive hepatitis B carriers and longer-duration therapy in patients with baseline HBV DNA levels $>2 \times 10^3$ IU/mL, until standard clinical endpoints are met (*Table 341-4*). Such chemotherapy-associated reactivation of hepatitis B is common (4–68%, median 25%, in a meta-analysis) in persons with ongoing HBV infection (HBsAg-reactive); however, such reactivation can occur, albeit less commonly, in persons who have cleared HBsAg but express anti-HBc (moderate risk, <10%) and rarely (<5%) even in persons with serologic evidence of recovery from HBV infection (anti-HBs-reactive, anti-HBc-reactive). Therefore, most authorities (e.g., Centers for Disease Control and Prevention; AASLD; American Gastroenterological Association; EASL) recommend HBsAg and anti-HBc (\pm anti-HBs) screening of all patients undergoing such chemotherapy and preemptive antiviral prophylaxis for HBsAg-reactive persons and anti-HBc-reactive persons treated with the most potent immunomodulatory agents (especially B cell-depleting agents like rituximab) and close on-therapy monitoring of other anti-HBc-reactive/anti-HBs-reactive persons with treatment if and when reactivation occurs.

CHRONIC HEPATITIS D (DELTA HEPATITIS)

Chronic hepatitis D virus (HDV) may follow acute co-infection with HBV but at a rate no higher than the rate of chronicity of acute hepatitis B. That is, although HDV co-infection can increase the severity of acute hepatitis B, HDV does not increase the likelihood of progression to chronic hepatitis B. When, however, HDV superinfection occurs in a person who is already chronically infected with HBV, long-term HDV infection is the rule, and a worsening of the liver disease is the expected consequence. Except for severity, chronic hepatitis B plus D has similar clinical and laboratory features to those seen in chronic hepatitis B alone. Relatively severe and progressive chronic hepatitis, with or without cirrhosis, is the rule, and mild chronic hepatitis is the exception. Occasionally, however, mild hepatitis or even, rarely, inactive carriage occurs in patients with chronic hepatitis B plus D, and the disease may

become indolent after several years of infection. A distinguishing serologic feature of chronic hepatitis D is the presence in the circulation of antibodies to liver-kidney microsomes (anti-LKM); however, the anti-LKM seen in hepatitis D, anti-LKM3, are directed against uridine diphosphate glucuronosyltransferase and are distinct from anti-LKM1 seen in patients with autoimmune hepatitis and in a subset of patients with chronic hepatitis C (see below). The clinical and laboratory features of chronic HDV infection are summarized in Chap. 339.

TREATMENT

Chronic Hepatitis D

Management is not well defined, and the host cellular RNA polymerase upon which HDV replication depends cannot be targeted by conventional antiviral agents. Glucocorticoids are ineffective and are not used. Preliminary experimental trials of IFN- α suggested that conventional doses and durations of therapy lower levels of HDV RNA and aminotransferase activity only transiently during treatment but have no impact on the natural history of the disease. In contrast, high-dose IFN- α (9 million units three times a week) for 12 months was reported to be associated with a sustained loss of HDV replication and clinical improvement in up to 50% of patients. Moreover, in anecdotal reports, the beneficial impact of treatment has been observed to persist for 15 years and to be associated with a reduction in grade of hepatic necrosis and inflammation, reversion of advanced fibrosis (improved stage), and clearance of HDV RNA in some patients. A suggested approach to therapy has been high-dose, long-term IFN for at least a year and, in responders, extension of therapy until HDV RNA and HBsAg clearance; however, extension of therapy to a second year provided no advantage, and sustained responses after completion of therapy have been rare. While standard IFN- α is the only approved drug for hepatitis D, PEG IFN has been shown to be more effective but still of limited therapeutic value; after 48 weeks of therapy, durable undetectable HDV RNA for 24 posttreatment weeks has been reported in a quarter to just over a half of patients. Disappointingly, loss of virologic responses (reappearance of HDV RNA) was observed during long-term (median 4.5 years) monitoring in a majority of 24-week-posttreatment responders, with durable HDV RNA suppression to undetectable in only 12%. Even extending PEG IFN therapy for 5 years and driving treatment doses up to 270 μ g weekly (of PEG IFN- α 2a), as reported in a small trial among 13 patients, while achieving serologic, virologic, histologic, biochemical, and clinical improvement, yielded sustained virologic responses (SVRs) in only 3 patients (58–246 weeks of posttreatment observation). None of the nucleoside analogue antiviral agents for hepatitis B is effective in hepatitis D, and adding oral nucleoside agents to PEG IFN is no more effective than PEG IFN monotherapy. While recommended, 12 months of PEG IFN therapy is far from satisfactory.

Preliminary trials have been performed with an oral prenylation inhibitor, lonafarnib, and with an inhibitor of HBV/HDV viral entry into hepatocytes, myrcludex B. Prenylation, the posttranslational covalent addition of the prenyl lipid farnesyl to large HDV antigen, is required for this HDV protein to interact and form secreted viral particles with HBsAg. In 14 patients treated twice daily for 28 days with 100 or 200 mg of lonafarnib, HDV RNA fell by $0.73 \log_{10}$ IU/mL and $1.54 \log_{10}$ IU/mL, respectively, before rebounding after completion of therapy. HBV entry into hepatocytes requires the binding of the myristolated N-terminal pre-S1 peptide of large HBsAg to sodium taurocholate co-transporting peptide, the functional receptor for HBV into hepatocytes. The application of myrcludex B, a synthetic homologous myristolated lipopeptide that competes for binding with HBsAg, was reported in a study of 24 patients (with a baseline mean of $4.1\text{--}4.2 \log_{10}$ copies/mL of HDV RNA) randomized to 24 weeks of treatment with myrcludex B (2 mg daily subcutaneously) as monotherapy or combined with PEG IFN compared to PEG IFN alone. A reduction

in HDV RNA occurred in all three groups, by $1.67 \log_{10}$ copies/mL (in two of eight patients RNA became undetectable), $2.59 \log_{10}$ copies/mL (in five of eight patients RNA became undetectable), and $2.17 \log_{10}$ copies/mL (in two of eight patients RNA became undetectable), respectively. No change occurred, however, in the level of HBsAg, which would have been expected. In these two exploratory brief-duration trials, sustained responses were not achieved, and toxicities were encountered (e.g., intermittent vomiting and weight loss [lonafarnib] and transient amylase and lipase elevations [myrcludex B]); however, from these proof-of-principle trials, more definitive and larger-scale studies of efficacy are awaited. Additional experimental approaches to the treatment of hepatitis D include nucleic acid polymer therapy to inhibit HBsAg release, administered alone or with PEG IFN and/or nucleoside analogues; so far, these studies have been done in one Eastern European site, have yielded some promising reductions in HDV RNA and HBsAg, but have been plagued by adverse effects, including marked ALT elevations. PEG IFN lambda has also been studied in small numbers of patients with hepatitis D; both IFN-associated side effects and elevations of aminotransferase and bilirubin levels accompanied modest on-treatment reductions in HDV RNA. Follow-up studies in larger numbers of patients have been frustratingly slow to materialize.

In patients with end-stage liver disease secondary to chronic hepatitis D, liver transplantation has been effective. If hepatitis D recurs in the new liver without the expression of hepatitis B (an unusual serologic profile in immunocompetent persons but common in transplant patients), liver injury is limited. In fact, the outcome of transplantation for chronic hepatitis D is superior to that for chronic hepatitis B; in such patients, combination HBIG and nucleoside analogue therapy for hepatitis B is indicated (Chap. 345).

CHRONIC HEPATITIS C

Regardless of the epidemiologic mode of acquisition of hepatitis C virus (HCV) infection, chronic hepatitis follows acute hepatitis C in 50–70% of cases; chronic infection is common even in those with a return to normal in aminotransferase levels after acute hepatitis C, adding up to an 85% likelihood of chronic HCV infection after acute hepatitis C. Few clues had emerged to explain host differences associated with chronic infection until recently, when variation in a single nucleotide polymorphism (SNP) on chromosome 19, *IL28B* (which codes for IFN- λ 3, now renamed *IFNL3*), was identified that distinguished between responders and nonresponders to IFN-based antiviral therapy (see below). The same variants correlated with spontaneous resolution after acute infection: 53% in genotype C/C, 30% in genotype C/T, but only 23% in genotype T/T. The association with HCV clearance after acute infection is even stronger when *IL28B* (*IFNL3*) haplotype is combined with haplotype G/G of a SNP near human leukocyte antigen (HLA) class II *DBQ1 03:01*.

In patients with chronic hepatitis C followed for 20 years, progression to cirrhosis occurs in ~20–25%. Such is the case even for patients with relatively clinically mild chronic hepatitis, including those without symptoms, with only modest elevations of aminotransferase activity, and with mild chronic hepatitis on liver biopsy. Even in cohorts of well-compensated patients with chronic hepatitis C referred for clinical research trials (no complications of chronic liver disease and with normal hepatic synthetic function), the prevalence of cirrhosis may be as high as 50%. Most cases of hepatitis C are identified initially in asymptomatic patients who have no history of acute hepatitis C (e.g., those discovered while attempting to donate blood, while undergoing lab testing as part of an application for life insurance, or as a result of routine laboratory tests). The source of HCV infection in many of these cases is not defined, although a long-forgotten percutaneous exposure (e.g., injection drug use) in the remote past can be elicited in a substantial proportion and probably accounts for most infections; most of these infections were acquired in the 1960s and 1970s among persons in the 1945–1965 birth cohort (Chap. 339), coming to clinical attention decades later.

Approximately one-third of patients with chronic hepatitis C have normal or near-normal aminotransferase activity; although one-third to one-half of these patients have chronic hepatitis on liver biopsy, the grade of liver injury and stage of fibrosis tend to be mild in the vast majority. In some cases, more severe liver injury has been reported—even, rarely, cirrhosis, most likely the result of previous histologic activity. Among patients with persistent normal aminotransferase activity sustained over ≥ 10 years, histologic progression has been shown to be rare; however, approximately one-fourth of patients with normal aminotransferase activity experience subsequent aminotransferase elevations, and histologic injury can be progressive once abnormal biochemical activity resumes. Therefore, continued clinical monitoring and antiviral therapy are indicated, even for patients with normal aminotransferase activity.

Despite this substantial rate of progression of chronic hepatitis C and even though liver failure can result from end-stage chronic hepatitis C, the long-term prognosis over 1–2 decades for chronic hepatitis C in most patients is relatively benign. Mortality over 10–20 years among patients with transfusion-associated chronic hepatitis C has been shown not to differ from mortality in a matched population of transfused patients in whom hepatitis C did not develop. Although death in the hepatitis group is more likely to result from liver failure and although hepatic decompensation may occur in ~15% of such patients over the course of a decade, the majority (almost 60%) of patients remain asymptomatic and well compensated, with no clinical sequelae of chronic liver disease. Overall, chronic hepatitis C tends to be very slowly and insidiously progressive, if at all, in most patients, whereas in approximately one-fourth of cases, chronic hepatitis C will progress eventually to end-stage cirrhosis. In fact, because HCV infection is so prevalent, and because a proportion of patients progress inexorably to end-stage liver disease, hepatitis C was the most frequent indication for liver transplantation ([Chap. 345](#)) in the era prior to the availability of direct-acting antiviral (DAA) therapy (see below). In the United States, hepatitis C accounts for up to 40% of all chronic liver disease; as of 2007, mortality caused by hepatitis C surpassed that associated with HIV/AIDS, and as of 2012, reported deaths caused by hepatitis C surpassed those associated with all other notifiable infectious diseases (HIV, tuberculosis, hepatitis B, and 57 other infectious diseases). Moreover, because the prevalence of HCV infection is so much higher in the “baby boomer” cohort born between 1945 and 1965, three-quarters of the mortality associated with hepatitis C occurs in this age cohort. Referral bias may account for the more severe outcomes described in cohorts of patients reported from tertiary care centers (20-year progression of $\geq 20\%$) versus the more benign outcomes in cohorts of patients monitored from initial blood-product-associated acute hepatitis or identified in community settings (20-year progression of only 4–7%). Still unexplained, however, are the wide ranges in reported progression to cirrhosis, from 2% over 17 years (eventually 19% over 36 years) in a population of Irish women with hepatitis C infection acquired from contaminated anti-D immune globulin to 30% over ≤ 11 years in recipients of contaminated intravenous immune globulin.

Progression of liver disease in patients with chronic hepatitis C has been reported to be more likely in patients with older age, longer duration of infection, advanced histologic stage and grade, more complex HCV quasispecies diversity, increased hepatic iron, concomitant other liver disorders (alcoholic liver disease, chronic hepatitis B, hemochromatosis, α_1 -antitrypsin deficiency, and steatohepatitis), HIV infection, and obesity. Among these variables, however, duration of infection appears to be one of the most important, and some of the others probably reflect disease duration to some extent (e.g., quasispecies diversity, hepatic iron accumulation). No other epidemiologic or clinical features of chronic hepatitis C (e.g., severity of acute hepatitis, level of aminotransferase activity, level of HCV RNA, presence or absence of jaundice during acute hepatitis) are predictive of eventual outcome. Despite the relatively benign nature of chronic hepatitis C over time in many patients, cirrhosis following chronic hepatitis C has been associated with the late development, after several decades, of HCC ([Chap. 82](#)); the annual rate of HCC in cirrhotic patients with hepatitis C is 1–4%, occurring primarily in patients who have had HCV infection for 30 years or more.

Perhaps the best prognostic indicator in chronic hepatitis C is liver histology; the rate of hepatic fibrosis may be slow, moderate, or rapid. Patients with mild necrosis and inflammation as well as those with limited fibrosis have an excellent prognosis and limited progression to cirrhosis. In contrast, among patients with moderate to severe necroinflammatory activity or fibrosis, including septal or bridging fibrosis, progression to cirrhosis is highly likely over the course of 10–20 years. The pace of fibrosis progression may be accelerated by such factors as concomitant HIV infection, other causes of liver disease, excessive alcohol use, and hepatic steatosis. Among patients with compensated cirrhosis associated with hepatitis C, the 10-year survival rate is close to 80%; mortality occurs at a rate of 2–6% per year; decompensation at a rate of 4–5% per year; and, as noted above, HCC at a rate of 1–4% per year. Estimates of the natural history of chronic hepatitis C have been made, based on data available on the prevalence of HCV infection in the U.S. population and on the rate of disease progression. Weighted primarily by the concentration of chronic hepatitis C in the baby boomer generation, the peak prevalence was estimated to have occurred in 2015. The calculated frequency of cirrhosis in U.S. patients with hepatitis C was 5% in 1990 and 25% in 2010 and was projected to be 37% in 2020. Estimated peak mortality has been predicted to occur in 2032. [A discussion of the pathogenesis of liver injury in patients with chronic hepatitis C appears in Chap. 339.](#)

Clinical features of chronic hepatitis C are similar to those described above for chronic hepatitis B. Generally, fatigue is the most common symptom; jaundice is rare. Immune complex-mediated extrahepatic complications of chronic hepatitis C are less common than in chronic hepatitis B (despite the fact that assays for immune complexes are often positive in patients with chronic hepatitis C), with the exception of essential mixed cryoglobulinemia ([Chap. 339](#)), which is linked to cutaneous vasculitis and membranoproliferative glomerulonephritis as well as lymphoproliferative disorders such as B-cell lymphoma and unexplained monoclonal gammopathy. In addition, chronic hepatitis C has been associated with extrahepatic complications unrelated to immune-complex injury. These include Sjögren’s syndrome, lichen planus, porphyria cutanea tarda, renal injury, type 2 diabetes mellitus, and the metabolic syndrome (including insulin resistance and steatohepatitis). In addition, a link has been observed between HCV infection and cardiovascular/cerebrovascular disease, rheumatologic/immunologic disorders, mental health and cognitive disorders, and nonliver malignancies.

Laboratory features of chronic hepatitis C are similar to those in patients with chronic hepatitis B, but aminotransferase levels tend to fluctuate more (the characteristic episodic pattern of aminotransferase activity) and to be lower, especially in patients with long-standing disease. An interesting and occasionally confusing finding in patients with chronic hepatitis C is the presence of autoantibodies. Rarely, patients with autoimmune hepatitis (see below) and hyperglobulinemia have false-positive immunoassays for anti-HCV. On the other hand, some patients with serologically confirmable chronic hepatitis C have circulating anti-LKM. These antibodies are anti-LKM1, as seen in patients with autoimmune hepatitis type 2 (see below), and are directed against a 33-amino-acid sequence of cytochrome P450 IID6. The occurrence of anti-LKM1 in some patients with chronic hepatitis C may result from the partial sequence homology between the epitope recognized by anti-LKM1 and two segments of the HCV polyprotein. In addition, the presence of this autoantibody in some patients with chronic hepatitis C suggests that autoimmunity may be playing a role in the pathogenesis of chronic hepatitis C.

Histopathologic features of chronic hepatitis C, especially those that distinguish hepatitis C from hepatitis B, are described in [Chap. 339](#).

TREATMENT

Chronic Hepatitis C

Therapy for chronic hepatitis C has evolved substantially in the 30 years since IFN- α was introduced for this indication in 1991. The therapeutic armamentarium grew to include PEG IFN with ribavirin and, then, in 2011, the introduction of the first protease

inhibitors, telaprevir and boceprevir, used in combination with PEG IFN and ribavirin in patients with HCV genotype 1. The field of antiviral therapy for hepatitis C was transformed beginning in 2013, with the approval of the first nucleoside analogue polymerase inhibitor, sofosbuvir. Although several of these combination regimens have been supplanted by better, later-generation drugs, as of 2020, five all-oral, highly effective (>95%), low-resistance, pan-genotypic, well-tolerated, short-duration (primarily 8–12 weeks) combination regimens of DAA drugs are recommended. The remarkable historical evolution of antiviral therapy for hepatitis C is instructive.

THE INTERFERON ERA (1991–2011)

IFN-based therapy has been supplanted by DAA agents introduced in the second decade of the twenty-first century; however, many important lessons about antiviral therapy for chronic hepatitis C were learned from the experience with IFN-based treatment, and many of the limitations of—and disparities in responsiveness to—IFN-based therapy have been overcome by current-generation DAA treatments. Mechanistically, HCV proteins inhibit several steps of the JAK-STAT signal transduction pathway, and by activation of JAK-STAT signaling, exogenous IFN culminates in and restores intracellular expression of IFN-stimulated genes and their protein products that have antiviral properties.

When first approved, subcutaneous IFN- α three times a week for 6 months achieved an SVR (Fig. 341-2) (defined then as a reduction of HCV RNA to PCR-undetectable levels ≥ 24 weeks after completion of therapy) in <10%. Doubling the duration of therapy increased the SVR rate to ~20%, and addition to the regimen of daily ribavirin (ineffective when used alone), an oral guanosine nucleoside, increased the SVR rate to 40% by reducing the likelihood of virologic relapse after completion of treatment (Fig. 341-2). Although its mechanism of action remains poorly understood, ribavirin retains a limited role in supporting DAA agents in several subgroups of otherwise refractory patients (see below).

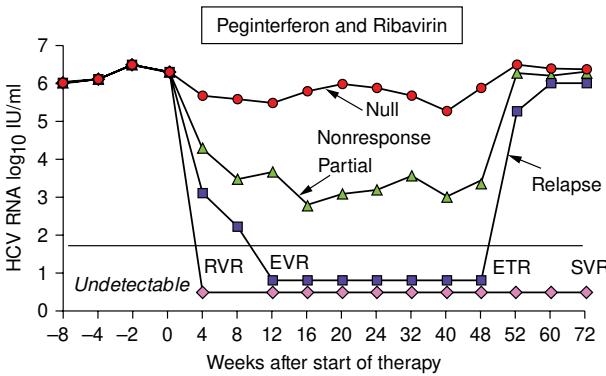


FIGURE 341-2 Classification of virologic responses based on outcomes during and after a 48-week course of pegylated interferon (PEG IFN) plus ribavirin antiviral therapy in patients with hepatitis C, genotype 1 or 4 (for genotype 2 or 3, the course would be 24 weeks). Nonresponders can be classified as null responders (hepatitis C virus [HCV] RNA reduction of $< 2 \log_{10}$ IU/mL) or partial responders (HCV RNA reduction $\geq 2 \log_{10}$ IU/mL but not suppressed to undetectable) by week 24 of therapy. In responders, HCV RNA can become undetectable, as shown with sensitive amplification assays, within 4 weeks (rapid virologic response [RVR]); can be reduced by $\geq 2 \log_{10}$ IU/mL within 12 weeks (early virologic response [EVR]); if HCV RNA is undetectable at 12 weeks, the designation is “complete” EVR; or can be undetectable at the end of therapy, 48 weeks (end-treatment response [ETR]). In responders, if HCV RNA remains undetectable for 24 weeks after ETR, week 72, the patient has a sustained virologic response (SVR), but if HCV RNA becomes detectable again, the patient is considered to have relapsed. The posttreatment week-24 SVR (SVR_{24}) has been supplanted by an SVR at week 12 (SVR_{12}), which has been shown to be equivalent to an SVR_{24} . In patients treated with direct-acting antiviral therapy, RVR and EVR milestones are largely irrelevant, being met by almost all patients. (Reproduced with permission from Marc G. Ghany, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health and the American Association for the Study of Liver Diseases. *Hepatology* 49:1335, 2009.)

Treatment with the combination of PEG IFN and ribavirin increased SVR rates to 55% overall—to >40% in genotypes 1 and 4, requiring 48 weeks of therapy, and to >80% in genotypes 2 and 3, requiring only 24 weeks of therapy—and histologic improvement in approximately three-fourths of patients. After initiation of IFN treatment, ALT levels fell precipitously, and up to 90% of virologic responses were achieved within the first 12 weeks of therapy. Failure to achieve an early virologic response (EVR), a $\geq \log_{10}$ reduction in HCV RNA by week 12, predicted failure to experience a subsequent SVR. Similarly, patients in whom HCV RNA became undetectable within 4 weeks (i.e., who achieve a rapid virologic response [RVR]) had a very high likelihood of achieving an SVR (Fig. 341-2). Surprisingly, however, high-dose induction with IFN-based therapy did not yield higher SVR rates.

Most relapses occurred within the first 12 weeks after treatment, and absence of HCV RNA 12 weeks after completion of therapy has become the current standard for SVR (SVR_{12}); relapses are very rare 6 months to a year after SVR and almost unheard of after 2 years. Of documented durability decades after successful therapy, an SVR to antiviral therapy for chronic hepatitis C is tantamount to a cure and is followed by marked improvements in liver-disease outcomes (see below).

Patient variables that correlated with IFN-based SVRs included favorable genotype (genotypes 2 and 3 as opposed to genotypes 1 and 4; genotype 1b as opposed to genotype 1a); low baseline HCV RNA level (<800,000 IU/mL), low HCV quasispecies diversity, and histologically mild hepatitis and minimal fibrosis, especially absence of cirrhosis; immunocompetence; low liver iron levels; age <40; female gender; and absence of obesity, insulin resistance, type 2 diabetes mellitus, and hepatic steatosis. High levels of HCV RNA, more histologically advanced liver disease, and high HCV quasispecies diversity all went hand in hand with advanced duration of infection and reduced IFN responsiveness. Also associated with poor responses to IFN-based therapy were African-American ethnicity (contributed to, but not explained entirely by, a higher proportion with genotype 1, slower early treatment viral kinetics, impaired HCV-specific immunity, and host genetic differences in *IL28B* [*IFNL3*] alleles, described below), Latino ethnicity, and poor treatment adherence (<80% of IFN and ribavirin doses and <80% of prescribed duration of therapy). Ironically, patients whose disease was least likely to progress were the ones *most* likely to respond to IFN and vice versa.

As described above in the discussion of spontaneous recovery from acute hepatitis C, IFN gene variants discovered in genome-wide association studies were shown to have a substantial impact on IFN responsiveness of patients with genotype 1 to antiviral therapy. In studies of patients treated with PEG IFN and ribavirin, variants of the *IL28B* (now renamed *IFNL3*) SNP that code for IFN- $\lambda 3$ (a type III IFN, the receptors for which are more discretely distributed than IFN- α receptors and more concentrated in hepatocytes) correlated significantly with responsiveness. Homozygotes for the C allele at this locus (C/C) achieved SVRs of ~80%, heterozygotes (C/T) SVRs of ~35%, and homozygotes for the T allele (T/T) SVRs of ~25%.

Side effects of IFN therapy are described in the section on treatment of chronic hepatitis B (see above). Besides ribavirin-associated nasal and chest congestion, pruritus, and precipitation of gout, the most pronounced ribavirin side effect is hemolysis, often requiring dose reduction or addition of erythropoietin therapy (not shown, however, to increase the likelihood of an SVR); therefore, close monitoring of blood counts is crucial, and ribavirin should be avoided in patients with anemia, hemoglobinopathies, coronary artery disease or cerebrovascular disease, or renal insufficiency (the drug is excreted renally) and in pregnancy (the drug is teratogenic, mandating contraception during, and for several months after, therapy in women of child-bearing age [because of their antiproliferative properties, IFNs also are contraindicated during pregnancy]). Overall, combination IFN-ribavirin therapy was more difficult to tolerate than IFN monotherapy and more likely to lead to dose reductions and discontinuation of therapy.

Beginning in 2011, for the treatment of hepatitis C, standard IFNs were supplanted by PEG IFNs, which have substantially longer half-lives, permitting administration once (rather than three times) a week. Once-a-week PEG IFN monotherapy was twice as effective as IFN monotherapy, approached the efficacy of combination standard IFN plus ribavirin, and was as well tolerated as standard IFNs. For most of the decade prior to 2011, when protease inhibitors were introduced for HCV genotype 1 (see below), the standard of care was a combination of PEG IFN plus ribavirin for all HCV genotypes.

Two PEG IFNs were available: PEG IFN- α 2b, a 12-kD, linear PEG molecule bound to IFN- α 2b, and PEG IFN- α 2a, a larger, 40-kD, branched PEG molecule bound to IFN- α 2a; because of its larger size and smaller volume of extravascular distribution, PEG IFN- α 2a could be given at a uniform dose independent of weight, whereas the dose of the smaller PEG IFN- α 2b, which has a much wider volume distribution, had to be weight-based. Between the two PEG IFNs, PEG IFN- α 2a appeared to be slightly better tolerated and slightly more effective than PEG IFN- α 2b in registration trials. The frequency of an SVR to PEG IFN–ribavirin therapy could be increased by tailoring therapy according to baseline variables and on-treatment virologic responsiveness. For example, in patients with baseline variables weighing against a response (e.g., high HCV RNA, obesity), raising the dose of PEG IFN and/or of ribavirin or extending therapy to 72 weeks for patients with genotype 1 and a slow virologic response, SVR rates could be improved. In contradistinction, in the \gtrsim 20% of patients with genotype 1 (and 4) who had a 4-week RVR and low baseline HCV RNA, treatment could be abbreviated to 24 weeks and SVR rates of \sim 90% achieved.

For most of the decade prior to 2011, when protease inhibitors were introduced for HCV genotype 1 (see below), the standard of care was a combination of PEG IFN plus ribavirin (unless ribavirin was contraindicated) for all HCV genotypes. Even after the introduction of protease inhibitors for genotypes 1 and 4, however, PEG IFN–ribavirin remained the standard of care for patients with genotypes 2 and 3 until late 2013. Responsiveness to IFN–ribavirin-based therapy was diminished in immunocompromised patients and in patients with HIV-HCV co-infection and contraindicated in patients with decompensated liver disease or end-stage renal disease. The cumbersome nature of IFN–ribavirin-based therapy (injections, complicated laboratory monitoring, side effects and poor tolerability, modest efficacy, variables and patient subsets associated with poor responsiveness, tailored therapy, futility rules, etc.) was supplanted eventually (in 2016) by DAAs for all genotypes (see below). Most of the variables associated with poor responsiveness to IFN-based therapy became irrelevant, and difficult-to-treat patient subpopulations began to experience responses to DAAs that were indistinguishable from responses in standard patients (see below).

Persons with chronic HCV infection suffer increased liver-related mortality, all-cause mortality, and multiple extrahepatic disorders (see above). On the other hand, successful antiviral therapy of chronic hepatitis C resulting in an SVR was shown to improve survival (and to reduce the need for liver transplantation); to lower the risk of liver failure, liver-related death, and all-cause mortality; to slow the progression of chronic hepatitis C; to reverse fibrosis and even cirrhosis; and to improve such HCV-associated extrahepatic disorders as type 2 diabetes and renal disease. Whereas the 10- and 20-year survival in the absence of an SVR is reduced in cirrhotic patients with chronic hepatitis C, survival at these intervals after an SVR has been found to be indistinguishable from that of the general population. In cirrhotic patients (and in those with advanced fibrosis), although successful treatment reduces mortality and liver failure (three- to fourfold 10-year reduction) and reduces the need for liver transplantation and the likelihood of HCC (14-fold 10-year reduction), the risk of liver-related death and HCC persists, albeit at a much reduced level, necessitating continued clinical monitoring and cancer surveillance after SVR in cirrhotics. On the other hand, in the absence of an SVR, IFN-based therapy does not reduce the

risk of HCC. Fortunately, PEG IFN–ribavirin nonresponders can now be retreated with DAAs and experience SVR rates comparable to those in treatment-naïve persons (see below).

FIRST-GENERATION PROTEASE INHIBITORS (2011–2013)

The HCV RNA genome encodes a single polyprotein, which is cleaved during and after translation by host and viral-encoded proteases. One protease involved in the cleavage of the viral polyprotein is an NS3/4A viral protein that has serine protease activity. Telaprevir and boceprevir are serine protease inhibitors that target NS3/4A. In 2011, telaprevir and boceprevir used in combination with PEG IFN and ribavirin were approved by the U.S. Food and Drug Administration (FDA) as the first oral DAA agents for the treatment of hepatitis C genotype 1 (not other genotypes) in adults with stable liver disease, both in patients who had not been treated before and in patients who had failed previous treatment. Although now replaced by more effective, all-oral regimens, these first-in-class agents represented a breakthrough in the treatment of chronic hepatitis C and established milestones against which subsequent therapies could be measured.

Because resistance developed rapidly during monotherapy with telaprevir and boceprevir, these drugs had to be used in combination with PEG IFN and ribavirin. Ribavirin in particular appeared to reduce relapse rates significantly in protease-inhibitor-based regimens, such that those who could not take or were intolerant to ribavirin were unlikely to benefit from the addition of these agents. Telaprevir and boceprevir regimens consisted of periods of triple therapy (protease inhibitor plus PEG IFN plus ribavirin) and periods of dual therapy (PEG IFN plus ribavirin). Telaprevir regimens began with 12 weeks of triple therapy followed by dual therapy of a duration based on HCV RNA status at weeks 4 and 12 (“response-guided therapy”) and prior treatment status. Boceprevir-based regimens consisted of a 4-week lead-in period of dual (PEG IFN–ribavirin) therapy followed by triple therapy and, in some instances, a further extension of dual therapy, with duration of response-guided therapy based on HCV RNA status at weeks 4, 8, and 24 and prior treatment status.

For patients with HCV genotype 1, protease inhibitors improved the frequency of RVRs and SVRs significantly as compared to PEG IFN plus ribavirin alone. In treatment-naïve patients, telaprevir-based SVRs were achieved in up to 79% of patients who received 12 weeks of triple therapy followed by 12–36 weeks of dual therapy, and among those with EVRs (undetectable HCV RNA at weeks 4 and 12) and response-guided therapy stopped at week 24 (12 weeks of triple therapy, then 12 weeks of dual therapy), SVRs occurred in 83–92%. In studies with boceprevir in treatment-naïve patients, SVRs occurred in 59–66% of patients, and among those with undetectable HCV RNA at 8 weeks, the SVR rate increased to 86–88%. Adding to the complexity of treatment with these protease inhibitors were absolute stopping rules for futility, that is, absence of HCV RNA reductions at critical treatment milestones, which were shown to be invariably predictive of nonresponse (telaprevir: HCV RNA >1000 IU/mL at weeks 4 or 12, or detectable at week 24; boceprevir: HCV RNA ≥ 100 IU/mL at week 12, or detectable at week 24).

In patients previously treated unsuccessfully with PEG IFN plus ribavirin, telaprevir-based treatment achieved SVRs in 83–88% of prior relapsers, 54–59% of partial responders (HCV RNA reduced by $\geq \log_{10}$ IU/mL but not to undetectable levels), and 29–33% of null responders (HCV RNA reduced by $<2 \log_{10}$ IU/mL). With boceprevir, a similar degradation in SVR rate occurred as a function of prior responsiveness—in 75% of prior relapsers, in 40–52% of previous partial responders, and in \sim 30–40% of null responders. In a substantial proportion of protease inhibitor nonresponders, resistance-associated substitutions (RASs, previously referred to as resistance-associated variants [RAVs]) could be identified, but these variants were not archived, and wild-type HCV reemerged in almost all cases within 1.5–2 years. SVRs to these protease inhibitors were highest in prior relapsers and treatment-naïve patients (white $>$ black ethnicity), lower in prior partial responders, lower still in

prior null responders, and lowest in cirrhotic prior null responders, for whom no benefit accrued over PEG IFN–ribavirin treatment. Responses to protease inhibitor triple-drug regimens were higher in patients with *IL28B* (*IFNL3*) C than non-C genotypes, HCV genotype 1b than genotype 1a, less advanced than more advanced fibrosis stage, whites than blacks, lower body mass index (BMI) than elevated BMI, and, for boceprevir, achievement of a $>1 \log_{10}$ HCV RNA reduction during 4 weeks of PEG IFN–ribavirin lead-in therapy. Age and HCV RNA level were less influential and insulin resistance was noninfluential on response to these antiviral agents.

Both protease inhibitors had substantial toxicities. Telaprevir was associated with a severe, generalized (trunk and extremities), often confluent, maculopapular, pruritic rash in ~6% of treated patients (that required careful dermatologic monitoring in all patients and systemic corticosteroid therapy in the most severely affected). Other common side effects included pruritus, rectal burning, nausea, diarrhea, fatigue, dysgeusia (altered or unpleasant taste), and anemia, which required close monitoring, could be relatively refractory, and occasionally required transfusion and even hospitalization (especially in cirrhotic prior nonresponders). Anemia occurred in half of boceprevir-treated patients, neutropenia in up to 30%, and thrombocytopenia in 3–4%. Other side effects of boceprevir included fatigue, nausea, headache, dysgeusia, dry mouth, vomiting, and diarrhea.

Both drugs came with an inconveniently high pill burden and had to be administered every 8 h with food (telaprevir with a 20-g fat meal). Use of protease inhibitors was further complicated by numerous drug-drug interactions. As telaprevir and boceprevir are both eliminated by and inhibit CYP3A4, these agents could not be administered with other medications that induce CYP3A4 or are dependent on CYP3A4 for elimination. Care had to be taken to examine for any potential interactions between these protease inhibitors and other medications the patient was taking, and a convenient website became available to check for such drug-drug interactions (www.hep-druginteractions.org).

Despite the improvement in SVRs with protease-inhibitor-based regimens for genotype 1 compared to PEG IFN–ribavirin (e.g., in treatment-naïve patients 66–79% vs 38–44%), triple-drug protease inhibitor therapy was hampered by amplified intolerance, the complexity of response-guided regimens and futility stopping rules, the inconvenience of thrice-daily dosing with meals and a high pill burden, the need for PEG IFN injections and ribavirin with all their intolerance, and multiple drug-drug interactions. Moreover, side effects appeared to be more severe and burdensome once these drugs entered practice, especially in cirrhotic nonresponders, in whom studies reported from Europe showed serious adverse events in up to 45% and deaths in up to 3%. All these issues, as well as rapidly accelerating progress on next-generation and all-oral DAA therapy (see below), conspired to temper enthusiasm for these new antivirals; after a brief stint as recommended therapy (2011–2013), these drugs became obsolete and are no longer recommended or available.

DIRECT-ACTING ANTIVIRAL COMBINATIONS OF SECOND-GENERATION PROTEASE INHIBITORS, FIRST-GENERATION POLYMERASE INHIBITORS, AND FIRST-GENERATION NS5A INHIBITORS (2014–2015)

Since late 2013, the number of new antiviral agents for hepatitis C has expanded substantially, and currently, PEG IFN-based treatments have been supplanted by five remaining therapeutic regimens, which are all oral, IFN-free, highly efficacious (>95% SVR), and well tolerated, with high barriers to resistance, simple dosing, low pill burdens, treatment durations as brief as 8–12 weeks, and pan-genotypic efficacy (Table 341-6). These drugs are distributed among three classes of DAAs: NS3/4 protease inhibitors (which cleave the single HCV polyprotein into constituent structural and nonstructural proteins [drug name ending in “-previr”]), NS5B nucleoside and nonnucleoside polymerase inhibitors (which interfere with the RNA-dependent RNA polymerase [a replicase]

involved in synthesis of viral RNA [drug name ending in “-buvir”]), and NS5A inhibitors (which interfere with a membrane-associated phosphoprotein essential to the HCV RNA replication complex [drug name ending in “-asvir”]).

The first of the new DAA agents (approved in November 2013) was simeprevir, a second-generation protease inhibitor for genotype 1, followed shortly thereafter (December 2013) by sofosbuvir, a pan-genotypic nucleoside polymerase inhibitor. For genotype 1, both of these agents had to be combined with PEG IFN and ribavirin; for genotypes 2 and 3, sofosbuvir was administered with ribavirin, without PEG IFN; however, these treatment regimens have been supplanted by combinations of all-oral, IFN-free DAAs, and ribavirin is rarely needed and retained only for very limited indications.

Simeprevir: When simeprevir was used with PEG IFN, its efficacy (genotype 1b > 1a) was similar to that of first-generation protease inhibitors but required only once-a-day dosing without the complexity of response-guided therapy. Similar to first-generation protease inhibitors, simeprevir was hampered by many drug-drug interactions and side effects (including photosensitivity, rash, and mild hyperbilirubinemia); moreover, patients, with HCV NS3 polymorphism Q80K had markedly reduced drug efficacy, necessitating pretreatment genetic testing and disqualifying approximately a third of potential treatment candidates. Little about simeprevir supported its adoption in combination with PEG IFN and ribavirin. On the other hand, the combination of simeprevir (150 mg) along with sofosbuvir (400 mg) for 12 weeks was found to be effective in treatment-naïve (97% SVR₁₂) or treatment-experienced (95% SVR₁₂) patients without cirrhosis and in treatment-naïve (88% SVR₁₂) or treatment-refractory (79% SVR₁₂) patients with cirrhosis. Like first-generation protease inhibitors, however, simeprevir was limited to genotype 1, required pretreatment genotyping that disqualified a third of recipients, usually required concomitant PEG IFN and ribavirin, had multiple drug-drug interactions and side effects, and was not competitive with the improved combinations that followed; therefore, simeprevir is no longer recommended.

Sofosbuvir: Sofosbuvir, the first nonprotease inhibitor DAA to be approved, has an excellent profile—high potency, high barrier to resistance, pan-genotypic activity, very well tolerated with limited adverse effects (most commonly mild fatigue, insomnia, headache, and nausea), once-daily oral administration, and relative freedom from major drug-drug interactions. Sofosbuvir has efficacy in all genotypes (1–6); in treatment-naïve subjects and prior nonresponders to PEG IFN-based and protease-inhibitor-based therapy; with PEG IFN–ribavirin or in IFN-free regimens; in combination with ribavirin or with NS5A inhibitors (see below); and for treatment periods as brief as 8–12 weeks. Currently, sofosbuvir is used in combination with one of two NS5A inhibitors and is a component of three of the five currently recommended DAA regimens (Table 341-6).

Sofosbuvir-ledipasvir: The DAA combination that has had a dominant role in the treatment of hepatitis C is sofosbuvir (400 mg) plus the NS5A inhibitor ledipasvir (90 mg) in a once-a-day, fixed-dose, single pill, approved in October 2014 for genotype 1 and in November 2015 for genotypes 4, 5, and 6. Phase 3 trials were conducted in treatment-naïve noncirrhotic patients, in treatment-naïve cirrhotic and noncirrhotic patients, and in treatment-experienced cirrhotic and noncirrhotic patients treated for 8, 12, or 24 weeks, both with and without ribavirin. In treatment-naïve noncirrhotics, an SVR₁₂ was achieved in 97–99% of subjects, and no benefit was observed by extending therapy from 12 to 24 weeks or by adding ribavirin. Moreover, for treatment-naïve, noncirrhotic patients with baseline HCV RNA $<6 \times 10^6$ IU/mL, a treatment duration of 8 weeks was as effective as one of 12 weeks (94–95% SVR₁₂), which may be a consideration for a proportion of patients. In cirrhotic patients, SVR₁₂ was achieved in 97–100% of treatment-naïve subjects (no advantage of extending therapy from 12 to 24 weeks or of adding ribavirin); however, for cirrhotic prior nonresponders to IFN-based therapy, 12 weeks of therapy was inferior (86% SVR₁₂)

TABLE 341-6 Indications and Recommendations for Antiviral Therapy of Chronic Hepatitis C^a

Standard Indications for Therapy		FAILED PRIOR PEG IFN/RIBAVIRIN THERAPY, NO CIRRHOSIS ^d
All patients with chronic HCV infection (detectable HCV RNA, with or without elevated ALT) except for those with short life expectancies owing to comorbid conditions.		<i>Genotype 1a and 1b</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks ledipasvir + sofosbuvir 12 weeks grazoprevir + elbasvir 12 weeks (without ELB NS5A RASs) <i>Genotype 2</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks <i>Genotype 3</i> sofosbuvir + velpatasvir 12 weeks (for patients without baseline NS5A RAS Y93H for velpatasvir) glecaprevir + pibrentasvir 16 weeks sofosbuvir + velpatasvir + voxilaprevir 12 weeks for patients with baseline NS5A RAS Y93H for velpatasvir <i>Genotype 4</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks grazoprevir + elbasvir 12 weeks (for prior relapse) ledipasvir + sofosbuvir 12 weeks <i>Genotypes 5, 6</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks ledipasvir + sofosbuvir 12 weeks
Any stage of fibrosis; pretreatment biopsy is no longer embraced and has been supplanted by noninvasive measures of fibrosis, e.g., imaging to determine liver elasticity.		
Responsiveness in groups previously refractory to interferon-based therapy (HIV-HCV co-infection, renal insufficiency, African-American and Latino ethnicity, <i>IL28B</i> non-C haplotype, obesity, insulin resistance, hepatic decompensation, etc.) is not diminished to contemporary direct-acting oral combination regimens.		
Retreatment Recommended		
Relapsers, partial responders, or nonresponders after a previous course of interferon-based therapy or prior direct-acting antiviral therapy (see genotype-specific recommendations below).		
Antiviral Therapy Not Recommended		
Pregnancy: No clinical studies of direct-acting antivirals during pregnancy are available. Ribavirin is contraindicated during pregnancy; therefore, any regimen including ribavirin should not be used. Sofosbuvir; sofosbuvir + ledipasvir; and paritaprevir-ritonavir + ombitasvir + dasabuvir are classified as pregnancy category B, but the other direct-acting antivirals do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment outweighs the potential for fetal risk.		
Therapeutic Regimens (based on AASLD-IDSA recommendations, www.hcvguidelines.org) ^b		
The European Association for the Study of the Liver (EASL) recommendations diverge slightly from AASLD-IDSA recommendations. ^c		
TREATMENT-NAÏVE OR RELAPSED AFTER PRIOR PEG IFN-RIBAVIRIN THERAPY		FAILED PRIOR PEG IFN-RIBAVIRIN THERAPY, COMPENSATED CIRRHOsis ^d
<i>Genotype 1a and 1b</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks ledipasvir + sofosbuvir 12 weeks (consider 8 weeks for noncirrhotic HIV-negative patients with HCV RNA <6 × 10 ⁶ IU/mL) grazoprevir + elbasvir 12 weeks (no cirrhosis or cirrhosis sans ELB NS5A RASs) <i>Genotype 2</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks <i>Genotype 3</i> sofosbuvir + velpatasvir 12 weeks (in cirrhotics, recommended only if without baseline NS5A RAS Y93H for velpatasvir) glecaprevir + pibrentasvir 8 weeks sofosbuvir + velpatasvir 12 weeks + weight-based ribavirin (in cirrhotics with baseline NS5A RAS Y93H for velpatasvir) sofosbuvir + velpatasvir + voxilaprevir 12 weeks (in cirrhotics with baseline NS5A RAS Y93H for velpatasvir) <i>Genotype 4</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks (12 weeks for HIV co-infection) ledipasvir + sofosbuvir 12 weeks (consider 8 weeks for noncirrhotic HIV-negative patients with HCV RNA <6 × 10 ⁶ IU/mL) grazoprevir + elbasvir 12 weeks <i>Genotypes 5, 6</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 8 weeks ledipasvir + sofosbuvir 12 weeks (except for genotype 6e)	<i>Genotype 1a</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 12 weeks grazoprevir + elbasvir 12 weeks (without ELB NS5A RASs) ledipasvir + sofosbuvir + RBV 12 weeks <i>Genotype 1b</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 12 weeks grazoprevir + elbasvir 12 weeks ledipasvir + sofosbuvir + RBV 12 weeks <i>Genotype 2</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 12 weeks <i>Genotype 3</i> sofosbuvir + velpatasvir + voxilaprevir 12 weeks glecaprevir + pibrentasvir 16 weeks grazoprevir + elbasvir 12 weeks sofosbuvir + velpatasvir + RBV 12 weeks <i>Genotype 4</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 12 weeks grazoprevir + elbasvir 12 weeks (for prior relapse) ledipasvir + sofosbuvir 12 weeks <i>Genotypes 5, 6</i> sofosbuvir + velpatasvir 12 weeks glecaprevir + pibrentasvir 12 weeks ledipasvir + sofosbuvir 12 weeks	
		FEATURES ASSOCIATED WITH REDUCED RESPONSIVENESS TO DIRECT-ACTING ANTIVIRAL COMBINATION THERAPY Genotype and subtype (genotype 1a less responsive than genotype 1b for several drugs) Treatment experience Advanced fibrosis (bridging fibrosis, cirrhosis) Reduced adherence

(Continued)

TABLE 341-6 Indications and Recommendations for Antiviral Therapy of Chronic Hepatitis C^a (Continued)

^aRapidly evolving new recommendations continue to be issued; for up-to-date treatment recommendations, please see www.hcvguidelines.org. ^bFor treatment-naïve patients, simplified treatment regimen recommendations are in bold font (based on broad applicability, pangenotypic coverage, and simplicity). For treatment-experienced patients, recommended regimens are in bold font, and alternative regimens are in standard font. ^cThe following EASL recommendations differ from those of AASLD-IDSA:

Genotype 1

For genotype 1a, noncirrhotic, prior IFN/RBV nonresponders, sofosbuvir + ledipasvir is not recommended.

For genotype 1b, treatment-naïve or -experienced patients, EASL retains paritaprevir/ritonavir + ombitasvir + dasabuvir for 12 weeks (for 8 weeks in patients with stage F0–F2 fibrosis).

For genotype 1b, noncirrhotic, treatment-naïve or -experienced patients with stage F0–F2 fibrosis, the recommended duration of grazoprevir + elbasvir is 8 weeks.

Genotype 3

For cirrhotic, treatment-naïve or -experienced patients (IFN-based regimen failures), sofosbuvir + velpatasvir is not recommended. For noncirrhotic patients with genotype 3, sofosbuvir + ledipasvir + voxilaprevir is not recommended.

Genotype 4

For genotype 4, prior IFN/RBV nonresponders, sofosbuvir + ledipasvir is not recommended. In treatment-naïve noncirrhotics, shorter duration (8 weeks) is not recommended for patients with HCV RNA $\leq 6 \times 10^6$ IU/mL.

^dFor nonresponders to prior direct-acting antiviral therapy (protease, polymerase, or NS5A inhibitors) and for decompensated cirrhosis, please consult www.hcvguidelines.org.

Abbreviations: AASLD, American Association for the Study of Liver Diseases; ALT, alanine aminotransferase; ELB NSSA RASs, elbasvir NS5A resistance-associated substitutions; HCV, hepatitis C virus; IFN, interferon; IDSA, Infectious Diseases Society of America; PEG IFN, pegylated interferon; IU, international units (1 IU/mL is equivalent to ~2.5 copies/mL); RASs, resistance-associated substitutions; RBV, ribavirin.

to 24 weeks of therapy (100% SVR₁₂). This combination, which is equally effective in patients with HIV-HCV co-infection and in African-American patients, has been shown to be highly effective in patients with decompensated cirrhosis and in patients with hepatitis C after liver transplantation and after kidney transplantation. Initially, sofosbuvir-ledipasvir was not recommended in patients with advanced renal failure; however, subsequently, the safety and efficacy of sofosbuvir-ledipasvir in patients with advanced renal failure were established, and the DAA was approved for this indication (November 2019). All sofosbuvir-containing regimens can be associated with severe bradycardia in patients taking the antiarrhythmic agent amiodarone, especially along with beta blockers; sofosbuvir-containing combinations are contraindicated with amiodarone. Drug-drug interactions are few, but P-glycoprotein inducers, such as St. John's wort and rifampin, and proton pump gastric acid inhibitors, such as omeprazole, may reduce sofosbuvir-ledipasvir concentrations. Generally, responsiveness to sofosbuvir-ledipasvir is not reduced in patients with baseline RASs to these agents, with the exception of treatment-experienced patients who have baseline NSSA RASs (see Table 341-6).

Paritaprevir-ritonavir, ombitasvir, and dasabuvir: The combination of ritonavir (100 mg)-boosted paritaprevir (150 mg), a protease inhibitor; ombitasvir (25 mg), an NSSA inhibitor; and dasabuvir (250 mg), a nonnucleoside polymerase inhibitor, with or without weight-based ribavirin (total of five drugs), was approved in December 2014 for genotypes 1 and 4. Paritaprevir-ritonavir and ombitasvir, formulated in a single tablet, are taken once daily, and both dasabuvir (a separate pill) and weight-based ribavirin (when included in the regimen) are taken twice daily. In clinical trials, this combination achieved SVR₁₂ rates of 87–100% in treatment-naïve and treatment-experienced patients with genotype 1; without ribavirin, this combination in genotype 1a was ~7% less responsive than in genotype 1b. Therefore, in treatment-naïve patients with genotype 1a, this combination was administered *with* ribavirin for 12 weeks in the absence of cirrhosis (95–97% SVR₁₂) or for 24 weeks in the presence of compensated cirrhosis (94% SVR₁₂), whereas in patients with genotype 1b, the combination did not require ribavirin, and the duration of therapy was 12 weeks for both noncirrhotics and cirrhotics (99–100% SVR₁₂). In prior nonresponders without cirrhosis, the combination was administered for 12 weeks, *with* ribavirin in genotype 1a (96% SVR₁₂) and *without* ribavirin in genotype 1b (100% SVR₁₂). In prior nonresponders with cirrhosis, the combination was administered for 24 weeks *with* ribavirin in genotype 1a (SVR₁₂ 100% in prior relapsers and partial responders, 95% in prior null responders [in whom treatment without ribavirin was associated with an 80% SVR₁₂]), but only for 12 weeks and *without* ribavirin in genotype 1b (100% SVR₁₂). For genotype 4, the regimen was given for 12 weeks with ribavirin but without dasabuvir in

treatment-naïve and treatment-experienced patients (100% SVR₁₂), including those with compensated cirrhosis. In July 2016, the FDA approved a long-acting formulation of dasabuvir, allowing once-a-day instead of twice-a-day treatment; for genotype 1a, twice-daily ribavirin dosing remained.

This combination was well tolerated with generally mild side effects, for example, fatigue, asthenia, insomnia, headache, and pruritus. Hyperbilirubinemia (primarily unconjugated) and elevations in alanine aminotransferase activity could occur but resolved during or shortly after treatment. Because of occasional hyperbilirubinemia and potential hepatotoxicity (FDA warning letter issued October 2015 regarding hepatic failure/decompensation reported in treated cirrhotic patients), this combination (and all subsequently introduced protease-inhibitor-containing combinations) was contraindicated in patients with decompensated cirrhosis, and treated cirrhotic patients had to be monitored closely for decompensation; however, the safety and efficacy of this combination was demonstrated for patients with advanced renal insufficiency. Similar to other regimens containing protease inhibitors, drug-drug interactions are common with other drugs that induce CYP3A4 or are dependent on CYP3A4 for elimination. Checking for potential drug-drug interactions was important prior to initiating therapy with this drug combination (www.hep-druginteractions.org). Responsiveness to this multidrug regimen was not reduced in patients with baseline RASs to these agents.

Compared to sofosbuvir-ledipasvir, this regimen had the disadvantage of requiring twice-a-day ribavirin therapy for genotype 1a and of being contraindicated in decompensated cirrhosis; however, it had the advantage of offering a 12-week, ribavirin-free regimen for prior null responders with cirrhosis and providing an option for patients with renal failure. Based on regimen simplicity and superiority, subsequent-generation, ribavirin-free combination DAAs have supplanted paritaprevir-ritonavir, ombitasvir, and dasabuvir; this regimen is no longer recommended at all by the AASLD; however, it is retained in EASL recommendations as an alternative regimen for genotype 1b only.

Sofosbuvir and daclatasvir: Daclatasvir, an NSSA inhibitor, along with the polymerase inhibitor sofosbuvir, was approved by the FDA in July 2015 for genotype 3 and in February 2016 for genotype 1. At the time of its approval for genotype 3, daclatasvir filled a need inadequately met by other available combination DAAs; however, eventually, recommendation of this combination regimen was extended to genotypes 1–4 in the United States and to all genotypes (1–6) in Europe. Daclatasvir, a 60-mg tablet, and sofosbuvir, a separate 400-mg tablet, were taken once a day for 12–24 weeks.

In clinical trials among treatment-naïve or treatment-experienced patients, SVR₁₂ rates for 12 weeks of daclatasvir plus sofosbuvir were 98% for genotype 1 (comparable results in genotypes 1a and 1b),

92% for genotype 2, and 89% for genotype 3. For noncirrhotic patients, the addition of ribavirin or the extension of therapy to 24 weeks did not improve efficacy. In patients with compensated cirrhosis, limited prospective data and data from observational cohorts suggested that extending therapy to 24 weeks, with or without ribavirin, improved efficacy. In cirrhotics, SVR₁₂ was achieved in 93% with Child-Pugh class A and B but in only 56% with class C decompensated cirrhosis. For patients with genotype 3 and cirrhosis, the combination was effective in treatment-naïve patients (94% SVR₁₂) but less so in prior nonresponders (69% SVR₁₂). Outcomes in patients with HIV-HCV co-infection were comparable.

Like other sofosbuvir-NS5A inhibitor combinations, daclatasvir plus sofosbuvir was well tolerated (mild fatigue, headache, nausea, or diarrhea in 5–14%) but could cause severe bradycardia when administered with amiodarone (contraindicated), especially along with beta blockers. Because daclatasvir is a substrate for CYP3A, CYP3A inducers can reduce daclatasvir levels, and CYP3A inhibitors reduce daclatasvir levels. Similarly, daclatasvir, an inhibitor of P-glycoprotein, OATP1B1 and OATP1B3, and breast cancer resistance protein (BCRP), can increase the levels of drugs that are substrates of these transporters. Responsiveness to daclatasvir-containing drug combination therapy was reduced in cirrhotic patients with genotype 1a and in both cirrhotic and noncirrhotic patients with genotype 3 who had baseline daclatasvir-associated NS5A RASs.

As new combination DAs were introduced, however, daclatasvir-sofosbuvir was less competitive and no longer filled a niche; it has been supplanted by better, later-generation combination DAs and is no longer recommended.

DIRECT-ACTING ANTIVIRAL COMBINATIONS OF THIRD-GENERATION PROTEASE INHIBITORS AND SECOND-GENERATION NS5A INHIBITORS (2016)

Elbasvir-grazoprevir: Elbasvir (50 mg), an NS5A inhibitor, combined in a single, fixed-dose pill with grazoprevir (100 mg), an NS3/4 protease inhibitor, was approved in January 2016 as a once-a-day (with or without food) treatment for genotypes 1 and 4. In clinical trials, a 12-week course was effective in treatment-naïve and treatment-experienced patients without cirrhosis or with compensated cirrhosis. In treatment-naïve patients, this combination yielded an SVR₁₂ in 92% of patients with genotype 1a, 99% with genotype 1b, and 100% with genotype 4 (very small numbers, however); 10 patients with genotype 6 were included, but only 80% achieved SVR₁₂. Cirrhotic and noncirrhotic patients had comparable rates of SVR₁₂, 97% and 94%, respectively. For this drug combination, however, ~11% of patients with genotype 1a harbor NS5A polymorphisms, that is, RASs, at baseline. If present, these NS5A RASs reduce efficacy of elbasvir-grazoprevir (unlike baseline RASs to most of the other combination DAA regimens described above and below) from 99 to 58% in treatment-naïve patients. Therefore, all patients with genotype 1a require baseline RAS testing; when these RASs were present, treatment extension to 16 weeks and the addition of weight-based ribavirin were documented to bring SVR₁₂ rates up to expected levels of close to 100%. In treatment-experienced patients, both extending treatment to 16 weeks and adding ribavirin were studied; however, generally, in the absence of baseline NS5A RASs, SVR₁₂ rates were not increased over those without ribavirin for 12 weeks (94–97%). For genotype 1a, among prior nonresponders to PEG IFN-ribavirin, 12 weeks of elbasvir-grazoprevir sufficed without ribavirin except for patients with baseline NS5A RASs, who required 16 weeks of therapy and ribavirin. Among nonresponders to prior protease inhibitor therapy, even in the absence of baseline NS5A RASs, ribavirin had to be added to a 12-week regimen; in the presence of baseline NS5A RASs, treatment was extended to 16 weeks and ribavirin added. For genotype 1b, NS5A RASs are not an issue, and the only subgroup requiring modification of a 12-week course of therapy were prior nonresponders to protease inhibitor regimens, for whom ribavirin was added. For genotype 4, the recommended regimen for all prior

nonresponders (whether to PEG IFN-ribavirin or protease inhibitor regimens) was 16 weeks of elbasvir-grazoprevir plus ribavirin. Now that simpler, improved combination regimens are available, for patients with NSSA RASs, extending the duration of elbasvir-grazoprevir and adding ribavirin have been abandoned (Table 341-6); however, *elbasvir-grazoprevir is one of the currently recommended DAA combinations* (Table 341-6).

This combination is just as effective in patients with HIV-HCV co-infection and in patients with advanced renal failure (including those requiring hemodialysis), but like all protease-inhibitor-including DAA combinations, it is contraindicated in decompensated cirrhosis. In this vein, like other protease inhibitor regimens, elbasvir-grazoprevir can be associated with aminotransferase elevations and potential hepatotoxicity; because these drugs are excreted by the liver, plasma drug concentrations may become elevated substantially in the presence of impaired hepatic function. Therefore, all treated patients should have ALT screening periodically during therapy, and the drug should be stopped for elevations exceeding tenfold or for elevations of conjugated bilirubin, alkaline phosphatase, or prothrombin time.

Elbasvir-grazoprevir is well tolerated, with only low levels of mild adverse effects (fatigue, headache, or nausea in 5–11%) seen just as frequently in placebo recipients. Both elbasvir and grazoprevir are substrates for CYP3A and are subject to multiple potential drug-drug interactions. Therefore, this combination should not be used with potent CYP3A inducers; conversely, CYP3A and OATP1B1 inhibitors can lead to untoward elevations of plasma elbasvir-grazoprevir concentrations. Checking for potential drug-drug interactions is advisable prior to initiating therapy (www.hep-druginteractions.org).

Compared to other available regimens for genotypes 1 and 4, elbasvir-grazoprevir has the disadvantage/inconvenience of requiring baseline NS5A RAS testing but the advantages of a comparable regimen for cirrhotics and noncirrhotics, for treatment-naïve and treatment-experienced patients, and for patients with normal renal function and with renal failure.

Sofosbuvir-velpatasvir: The combination in a single, fixed-dose pill of velpatasvir (100 mg), a highly potent, pangenotypic NS5A inhibitor, and the polymerase inhibitor sofosbuvir (400 mg) was approved in June 2016 for genotypes 1–6 in treatment-naïve and treatment-experienced noncirrhotics and cirrhotics. In August 2017, approval was extended to include patients with HCV-HIV co-infection. Ribavirin is not required, including in patients with genotypes 2 and 3, except in patients with decompensated cirrhosis.

In a series of clinical trials, this combination for 12 weeks in the absence of ribavirin was shown to yield a 99% SVR₁₂ (range 97–100%) in genotypes 1, 2, 4, 5, and 6 and 95% in genotype 3. Baseline NS5A RASs had no impact on responsiveness.

Prior to the availability of this drug combination, patients with genotype 3, especially those with cirrhosis and prior null response to other therapies, proved to be the most refractory subset of patients. In treatment-naïve patients with genotype 3, 12 weeks of sofosbuvir-velpatasvir (95% SVR₁₂) was superior to 24 weeks of sofosbuvir plus ribavirin (80% SVR₁₂). In patients with genotype 3, the combination of sofosbuvir-velpatasvir for 12 weeks was comparable in noncirrhotics (97% SVR₁₂) and cirrhotics (91% SVR₁₂) and in treatment-naïve (97% SVR₁₂) and treatment-experienced (90% SVR₁₂) patients and was superior in all these categories to 24 weeks of sofosbuvir plus ribavirin (87%, 66%, 86%, and 63%, respectively). In cirrhotic null responders, most available IFN-free regimens for genotype 3 (including daclatasvir plus sofosbuvir, which had been approved specifically for this genotype) achieved SVR₁₂ rates in the range of ~60–75%, while the combination of PEG IFN, ribavirin, and sofosbuvir could boost SVR₁₂ to the mid-80% range. For treatment-experienced patients with genotype 3, sofosbuvir-velpatasvir in noncirrhotics and cirrhotics had similarly high efficacy (91% and 89% SVR₁₂, respectively); this was the highest recorded SVR₁₂ for genotype 3 cirrhotic null responders treated with IFN-free DAA regimens. Finally, in patients with

genotypes 1–4 and 6 and with decompensated, class B cirrhosis (55% treatment-experienced), sofosbuvir–velpatasvir plus ribavirin for 12 weeks yielded an SVR₁₂ in 94%; this result was better than sofosbuvir–velpatasvir without ribavirin for 12 weeks (83% SVR₁₂) or 24 weeks (86% SVR₁₂).

Like other all-oral DAAs, sofosbuvir–velpatasvir was very well tolerated; in noncirrhotic and compensated cirrhotic patients, mild headache and fatigue were seen in >10% (this occurred in a comparable proportion of placebo recipients), and in patients with decompensated cirrhosis, mild fatigue, headache, nausea, insomnia, diarrhea, and anemia (ribavirin was part of the regimen) were seen in >10%. Like other sofosbuvir-containing regimens, sofosbuvir–velpatasvir should not be administered along with amiodarone (potential serious bradycardia); in addition, P-glycoprotein inducers and moderate-to-potent CYP3A inducers can reduce plasma levels of sofosbuvir and/or velpatasvir. Checking for drug-drug interactions prior to therapy is advisable (www.hep-druginteractions.org). Baseline RASs do not influence responsiveness to this combination. Sofosbuvir–velpatasvir is one of the currently recommended DAA combinations for hepatitis C (Table 341-6). *Because it is so simple and broadly effective across patient subgroups, sofosbuvir–velpatasvir is one of the two combination DAA regimens recommended by the AASLD and EASL as a preferred, simplified treatment algorithm* (Table 341-6).

DIRECT-ACTING ANTIVIRAL COMBINATIONS OF THIRD-GENERATION NS5A INHIBITORS AND FOURTH-GENERATION PROTEASE INHIBITORS—CURRENT STANDARD OF CARE (SINCE 2017)

Sofosbuvir–velpatasvir–voxilaprevir: Approved in July 2017, the pangenotypic, high-barrier-to-resistance protease inhibitor voxilaprevir (100 mg) added to the polymerase inhibitor–NS5A inhibitor combination of sofosbuvir–velpatasvir yields a very well-tolerated triple-drug combination with ~97% SVR₁₂ across all HCV genotypes and patient subgroups. These include the small percentage of patients with genotype 1 and genotype 3 refractory to previously approved DAA combinations as well as noncirrhotic/cirrhotic, treatment-naïve/treatment-experienced groups, including those who had or who had not received prior NS5A treatment. Efficacy was independent of the number of prior DAA drug classes received, and no effects of baseline NS5A RASs were noted.

The potential for abbreviated (8-week) treatment with this triple combination was explored in a clinical trial involving treatment-naïve patients; however, the shortened duration was inferior to a full 12-week course. The side effect profile for sofosbuvir–velpatasvir–voxilaprevir was similar to that in the placebo arm of clinical trial patients and included mild and uncommon headache, fatigue, nausea, and diarrhea.

Because other DAA regimens are so effective in most patients with chronic hepatitis C, recommendations for sofosbuvir–velpatasvir–voxilaprevir are limited to a small subset of otherwise refractory patients: for treatment-naïve cirrhotic patients with genotype 3 and baseline NS5A velpatasvir RAS Y93H, for treatment-naïve (according to AASLD, not EASL) or IFN-ribavirin–experienced noncirrhotic or cirrhosis patients with genotype 3 (Table 341-6), and for patients with or without compensated cirrhosis and prior, failed NS5A inhibitor-containing therapy (consult www.hcvguidelines.org).

This triple-drug combination, like all sofosbuvir-containing combinations, is contraindicated in patients taking amiodarone and, like all protease inhibitor-containing combinations, in patients with decompensated cirrhosis. Concomitant omeprazole, 20 mg, can be taken with this sofosbuvir-containing regimen. Prior to initiating therapy, checking for drug-drug interactions is recommended.

Glecaprevir–pibrentasvir: A regimen of 8 weeks of this single-pill, fixed-dose combination of the protease inhibitor glecaprevir (300 mg) and NS5A inhibitor pibrentasvir (120 mg), two pangenotypic, high-potency DAAs with high barriers to resistance (approved in August 2017), achieves SVR₁₂ in close to 100% of treatment-naïve patients with all genotypes, with or without cirrhosis: SVR₁₂ of

~99% for genotypes 1, 2, and 4–6 and of 95–98% for genotype 3. Extended treatment for 12 weeks did not increase efficacy. In trials among treatment-experienced patients, treatment with 12 weeks of this DAA combination was just as effective as 16 weeks for all genotypes except genotype 3; however, with increasing numbers of prior treatment courses, SVR₁₂ rates fell—100% for patients treated with a protease inhibitor only, 88% for patients treated with an NS5A inhibitor only, and 79% for patients treated previously with both a protease inhibitor and an NS5A inhibitor. Similarly, baseline RASs reduced SVR₁₂ rates—from 100% without RASs (or with RASs limited to those reflecting protease inhibitor resistance) to 89% for baseline NS5A RASs.

For retreatment of patients with prior glecaprevir–pibrentasvir failure, 16 weeks of glecaprevir–pibrentasvir plus sofosbuvir are recommended (alternatively, sofosbuvir–velpatasvir–voxilaprevir for 12 weeks [+ ribavirin in cirrhotics]). Glecaprevir–pibrentasvir for 16 weeks is recommended as well after failure to respond to the triple-drug combination of sofosbuvir–velpatasvir–voxilaprevir (see below). For retreatment of patients with sofosbuvir–velpatasvir–voxilaprevir failure, 16 weeks of glecaprevir–pibrentasvir plus ribavirin is recommended, as is a repeat course of sofosbuvir–velpatasvir–voxilaprevir plus ribavirin for 24 weeks.

As is the case for any DAA combination containing a protease inhibitor, glecaprevir–pibrentasvir is contraindicated in decompensated cirrhosis; it has been shown to achieve an SVR₁₂ in 98% of patients with stage 4 or 5 renal disease (in treatment-naïve or experienced, cirrhotic or noncirrhotic patients) and is a preferred treatment for patients with severe renal impairment. This DAA combination should be taken with food, and drug-drug interactions should be considered prior to initiating treatment. *Because it is so simple and broadly effective across patient subgroups (8 weeks for all noncirrhotic treatment-naïve patients except patients with HIV co-infection [12 weeks]; 12 weeks for all treatment-experienced cirrhotics and treatment-naïve cirrhotics with genotype 3 [except treatment-experienced cirrhotic or noncirrhotic genotype 3 (16 weeks)]), glecaprevir–pibrentasvir is one of the two combination DAA regimens recommended by the AASLD and EASL as a preferred, simplified treatment algorithm* (Table 341-6).

Emerging data on the impact of DAAs on the natural history of chronic hepatitis C indicated that, as was documented for IFN-based therapy, successful DAA therapy is associated with a gradual reduction in fibrosis progression and a regression of advanced fibrosis (cirrhosis), improvement in survival among patients with decompensated cirrhosis, a reduction in HCC, and a decline in the number of patients with hepatitis C being referred for liver transplantation. Early observations purported to show an *increase* in HCC after a DAA-associated SVR for chronic hepatitis C. On the contrary, HCC rates are reduced dramatically and consistently after successful DAA therapy. Ultimately, the initial observation was attributed to a cohort bias resulting from the application of simple-to-use DAA therapy to an older and sicker population with more advanced chronic hepatitis C (including decompensated cirrhosis); this cohort effect explains why the *baseline* risk for HCC was higher in DAA-treated patients than it had been when IFN-based therapy was withheld from such patients. Thus, the increased risk in HCC cases was not linked to DAA treatment but to more advanced liver disease at baseline in patients treated with DAAs. The reports of HCC after DAA therapy drive home the residual HCC risk after SVR in patients with cirrhosis (advanced hepatic fibrosis) treated either in the IFN or DAA era; therefore, continued HCC surveillance after therapy is recommended for anyone with baseline advanced fibrosis prior to therapy.

Based on the known prevalence, natural history, and rate of progression of chronic hepatitis C and on the efficacy of DAA therapies and their impact on the complications of hepatitis C, modeling estimates have suggested that the availability and application of these therapies have the potential to reduce the hepatitis C-associated disease burden, including liver-related death, HCC, decompensated cirrhosis, and liver transplantation, by 50–70% between 2015 and 2050.

Because the pace of new drug development and approval has been so rapid, the AASLD and the Infectious Diseases Society of America (IDSA) have been providing a consensus of updated treatment recommendations for patients with hepatitis C; these recommendations, which continue to be revised regularly based on new data, are available online at www.hcvguidelines.org and should be consulted before initiating therapy (Table 341-6). The EASL issues similar (but not identical) treatment recommendations annually for hepatitis C (www.easl.eu), most recently in November 2020. Divergences between AASLD-IDSA and EASL recommendations are noted in Table 341-6.

Prior to therapy, HCV genotype should be determined, because the genotype contributes to decisions about which treatment regimens are indicated (Table 341-6). Monitoring of serum HCV RNA levels before, during, and after treatment is crucial in assessing response to therapy; moreover, the baseline level may contribute to determining the duration of therapy (e.g., in noncirrhotic patients with genotype 1 and HCV RNA $<6 \times 10^5$ IU/mL, 8 [instead of the usual 12] weeks of sofosbuvir-ledipasvir may be a consideration). The goal of treatment is to eradicate HCV RNA during therapy and to document that the virus remains undetectable for at least 12 weeks after completion of therapy (SVR₁₂). Several reports have appeared describing hepatitis B reactivation, often severe, during and after DAA therapy in patients co-infected with HCV and HBV who were not being treated for their HBV infections. Therefore, screening for HBV infection is recommended prior to initiating DAA therapy for hepatitis C (which should have been done to determine HBV immunity status as a prelude to recommended hepatitis B vaccination in patients with chronic hepatitis C), and therapy for HBV infection (for those meeting HBV treatment criteria, see above) should be initiated prior to or simultaneously with HCV therapy.

Because of their high efficacy and pangenotypic range, two DAA regimens, glecaprevir-pibrentasvir (8 weeks) and sofosbuvir-velpatasvir (12 weeks), are recommended as simplified treatment algorithms that can be prescribed for all treatment-naïve patients with or without cirrhosis (Table 341-6).

INDICATIONS FOR ANTIVIRAL THERAPY

Patients with chronic hepatitis C who have detectable HCV RNA in serum, whether or not aminotransferase levels are increased, and chronic hepatitis of any grade and stage are candidates for antiviral therapy with DAA agents. The only exception would be patients with short life expectancies, for whom treating hepatitis C would have no influence on longevity. Certainly, for patients with advanced liver disease, early treatment merits a high priority. Although patients with persistently normal aminotransferase activity tend to progress histologically very slowly or not at all, they respond to antiviral therapy just as well as do patients with elevated aminotransferase levels; therefore, such patients are candidates for antiviral therapy. As noted above, antiviral therapy has been shown to improve survival and complication-free survival and to slow progression of and to reverse fibrosis.

HCV genotype determines the regimen to be selected (Table 341-6). Similarly, the absence or presence of cirrhosis or advanced fibrosis determines the treatment options from which to select, including the antiviral agents to be used, the duration of therapy, and the now rare need for ribavirin (Table 341-6). In the past, a pretreatment liver biopsy was relied upon to assess histologic grade and stage as well as to identify such histologic factors as steatosis, which can influence responsiveness to therapy. As therapy has improved for patients with a broad range of histologic severity and as noninvasive measures of the stage of fibrosis (e.g., assessment of liver elasticity by imaging, FIB-4 score [see above]) have gained in accuracy and popularity, noninvasive approaches have supplanted histology in almost most cases. As noted above, if cirrhosis or advanced fibrosis is present prior to therapy, the risk of HCC, although reduced substantially by successful therapy, is

not eliminated, and twice-yearly posttreatment imaging for HCC surveillance (and endoscopic surveillance for esophageal varices at intervals of 1–3 years) is indicated even after an SVR. In patients with low-level fibrosis at baseline, achievement of an SVR allows the cessation of such surveillance.

Patients who have relapsed after, or failed to respond to, a course of IFN-based or DAA agent-based therapy are candidates for retreatment with a DAA therapy regimen (Table 341-6). For patients who have failed to respond to a DAA combination, options include increasing the duration of therapy with the failed regimen, adding ribavirin, or changing the drug class (e.g., after failed protease and polymerase inhibitors, switching to an NSSA-containing combination). In the presence of cirrhosis or a need for urgent retreatment, patients who have failed protease inhibitor plus polymerase inhibitor combination therapy or who have failed an NSSA combination are candidates for RAS testing and tailored therapy based on such resistance testing. If reliable RAS testing is not available, adding ribavirin or extending the duration of therapy are options. For prior nonresponders to IFN-based therapy, NSSA inhibitor-containing regimens are highly effective; however, reduced responsiveness can be encountered, especially in cirrhotic patients. For this relatively refractory group, ideally, the most potent or effective NSSA regimen should be selected to give such patients the best chance of responding and to avoid treatment-emergent NSSA RASs. Noted above (see discussion of sofosbuvir-velpatasvir-voxilaprevir and of glecaprevir-pibrentasvir) are potential retreatment approaches after failure of a prior NSSA-containing regimen. Additional details for treatment of such patient subgroups can be found at www.hcvguidelines.org. It is worth reiterating that protease inhibitors are contraindicated for patients with decompensated cirrhosis, and sofosbuvir-containing regimens are not recommended for patients taking amiodarone (especially with beta blockers) for treatment of cardiac arrhythmias. While sofosbuvir-containing DAA combinations were not recommended initially for patients with advanced renal failure, subsequent studies demonstrated safety and efficacy in this subgroup, and sofosbuvir-containing DAA combinations are now approved for advanced renal failure.

Persons with acute hepatitis C are also candidates for antiviral therapy (Chap. 339) with the same pangenotypic combination DAA agents (and the same duration of treatment) approved for chronic hepatitis C; delaying the initiation of therapy to allow for spontaneous recovery is no longer recommended. According to EASL recommendations, patients with acute hepatitis C should be treated ideally with a currently recommended 8-week DAA regimen. In patients with biochemically and histologically mild chronic hepatitis C, the rate of progression is slow; however, such patients respond just as well to antiviral therapy as those with elevated aminotransferase levels and more histologically severe hepatitis. Because of the high cost of DAA treatments, in the past, a higher priority was assigned to patients with advanced fibrosis/cirrhosis; however, this controversial approach was relied upon by some medical insurers and pharmacy benefit management organizations to withhold therapy from patients with low-level fibrosis. Unfortunately, delaying therapy until fibrosis becomes advanced misses the opportunity to prevent all the dire consequences of chronic hepatitis C (liver failure, death/transplantation, HCC), which can be reduced, but not eliminated completely, once advanced fibrosis is established. Therefore, therapy for patients with mild disease is justified as well as cost-effective.

Patients with compensated cirrhosis can respond to therapy, and their likelihood of a sustained response with DAAs is comparable to that in noncirrhotics. Patients with decompensated cirrhosis, who were not candidates for IFN-based antiviral therapy, respond well to DAA therapy regimens consisting of combinations of polymerase inhibitors and NSSA inhibitors (e.g., sofosbuvir-ledipasvir, sofosbuvir-velpatasvir); however, protease-inhibitor-containing combinations have been associated with potential hepatotoxicity and hepatic decompensation and, as noted above, are contraindicated in this patient subset. For decompensated cirrhosis, ribavirin should be

added to a 12-week course of sofosbuvir-NS5A therapy; however, in cases of ribavirin ineligibility, the duration of therapy should be extended to 24 weeks. In cases of prior failure to respond to sofosbuvir-NS5A therapy, the sofosbuvir-NS5A regimen should be repeated but supplemented with ribavirin and extended to 24 weeks (www.hcvguidelines.org). Patients with decompensated cirrhosis should be referred to a liver transplantation center. DAAs are highly effective not only for patients with end-stage liver disease awaiting liver transplantation but also for patients with recurrent hepatitis C after liver transplantation. Ideally, patients should be treated prior to liver transplantation; however, a concern is that eradication of HCV infection will disqualify such patients from accepting donor livers from persons with HCV infection, thus contracting the potential donor pool and limiting accessibility to donor organs and timely transplantation. In addition, responsiveness to DAA therapy appears to be reduced in patients with decompensated cirrhosis and with high Model for End-Stage Liver Disease (MELD) scores; in this subgroup, responsiveness after liver transplantation would be substantially better. Therefore, advocacy has been expressed for postponing DAA therapy in patients with high-MELD-score ($\geq 18-20$), HCV-associated, end-stage liver disease until after liver transplantation; for patients with MELD scores $< 18-20$, pretransplantation DAA therapy is advised. Still, the decision whether to treat pretransplantation or posttransplantation should be individualized thoughtfully for each patient, based on such factors as MELD score, time anticipated prior to availability of a donor organ, relative clinical stability, and comorbidities (Chap. 345). Because DAA therapy is so effective, many transplantation centers, to expand the donor pool, are accepting organs from HCV-infected donors, transplanting them into HCV-uninfected recipients, and treating recipients with sofosbuvir-velpatasvir for 12 weeks or gecaprevir-pibrentasvir for 8 weeks after transplantation—with excellent results.

The cutaneous and renal vasculitis of HCV-associated essential mixed cryoglobulinemia (Chap. 339) may respond to antiviral therapy, but sustained responses were rare after discontinuation of therapy in the IFN era, and prolonged, potentially indefinite, therapy was recommended. Now that more effective DAAs are available, a 12-week course of sofosbuvir-based combination therapy has been shown to yield an SVR₁₂ rate exceeding 80% in cryoglobulinemic vasculitis. Anecdotal reports suggest that IFN-based antiviral therapy may be effective in porphyria cutanea tarda or lichen planus associated with hepatitis C; whether the more appealing DAAs are effective in these groups remains to be documented.

In patients with HCV/HIV co-infection, hepatitis C is more progressive and severe than in HCV-monoinfected patients. Although patients with HCV/HIV co-infection responded less well to IFN-based antiviral therapy for hepatitis C, they respond as well as patients with HCV infection alone to DAA combination regimens. For patients with HCV/HIV co-infection, an abbreviated, 8-week course of sofosbuvir-ledipasvir for low-level HCV RNA is not recommended, and a full 12 weeks should be given; similarly, for patients with genotype 4, a 12-week course of gecaprevir-pibrentasvir is recommended instead of an 8-week course for treatment-naïve or -experienced patients with or without cirrhosis (Table 341-6). In HCV/HIV-infected patients, ribavirin can potentiate the toxicity of didanosine (e.g., lactic acidosis) and the lipotrophy of stavudine, and zidovudine can exacerbate ribavirin-associated hemolytic anemia; therefore, these drug combinations should be avoided. In HCV/HIV co-infected persons, the list of potential drug-drug interactions is extensive and should be consulted carefully before beginning DAA treatment (www.hcvguidelines.org).

Patients with a history of injection drug use and alcoholism can be treated successfully for chronic hepatitis C, preferably in conjunction with drug and alcohol treatment programs. Moreover, because injection drug users, as a source of transmission to others, account disproportionately for perpetuating the spread of HCV infection in the population, the impact of treating active injection drug users is amplified by reducing such transmission.

The approved oral DAA combinations are effective in patients with mild-modest renal failure and require no dose adjustments. For patients with severe renal impairment (creatinine clearances < 30 mL/min), including those undergoing hemodialysis, recommended combinations are 12 weeks of elbasvir-grazoprevir for genotypes 1 and 4 or 12 weeks of gecaprevir-pibrentasvir for all genotypes. Both in severe renal impairment and after renal transplantation, levels of SVR₁₂ in patients treated with these oral DAA combinations have approached 100%. Initially, in patients with severe renal impairment, sofosbuvir-containing combinations were not recommended. Subsequently, however, based on efficacy and safety in a series of clinical trials, sofosbuvir-containing regimens were approved by the FDA in November 2019 for patients with severe renal impairment.

No clinical studies of the use of DAAs during pregnancy are available. Ribavirin is contraindicated during pregnancy; therefore, any regimen including ribavirin should not be used. Sofosbuvir; sofosbuvir-ledipasvir; and paritaprevir-ritonavir, ombitasvir, and dasabuvir are classified as pregnancy category B; the other DAAs do not have a pregnancy classification. Therefore, these therapies are not indicated routinely in pregnancy and should be used, with caution, only if the benefit of treatment is compelling and justified compared to the potential for fetal risk. Currently, screening all pregnant women for HCV infection is recommended. Breast feeding is not contraindicated in women with HCV infection (unless the mother has a break in the integrity of the nipples or is co-infected with HIV).

Choosing Among Available Treatment Options Choosing among the number of all-oral DAA combinations approved since 2013 was daunting to treating clinicians. Currently, however, the number of recommended DAA combinations has narrowed to a very manageable few. The most popular of the regimens have been fixed-dose, single-pill, pangenotypic combinations. Although sofosbuvir-ledipasvir and elbasvir-grazoprevir are among the recommended DAA combinations (Table 341-6), for simplicity, two “one-size-fits-all” pangenotypic regimens—sofosbuvir-velpatasvir and gecaprevir-pibrentasvir—can be used, for 8–12 weeks, mostly without ribavirin, in almost all treatment-naïve, noncirrhotic and cirrhotic patients, including those with advanced renal failure and HCV-HIV co-infection. Applicability of the triple-drug combination sofosbuvir-velpatasvir-voxilaprevir is quite limited in treatment-naïve patients, reserved primarily for cirrhotic patients with genotype 3. As noted above, protease-inhibitor-containing DAA regimens (elbasvir-grazoprevir, gecaprevir-pibrentasvir, and sofosbuvir-velpatasvir-voxilaprevir) are contraindicated in decompensated cirrhosis.

AUTOIMMUNE HEPATITIS

■ DEFINITION

Autoimmune hepatitis is a chronic disorder characterized by continuing hepatocellular necrosis and inflammation, usually with fibrosis, which can progress to cirrhosis and liver failure. When fulfilling criteria of severity, this type of chronic hepatitis, when untreated, may have a 6-month mortality of as high as 40%. Based on contemporary estimates of the natural history of autoimmune hepatitis, the 10-year survival is 80–98% for treated and 67% for untreated patients. The prominence of extrahepatic features of autoimmunity and seroimmunologic abnormalities in this disorder supports an autoimmune process in its pathogenesis; this concept is reflected in the prior labels *lupoid* and *plasma cell hepatitis*. Autoantibodies and other typical features of autoimmunity, however, do not occur in all cases; among the broader categories of “idiopathic” or cryptogenic chronic hepatitis, many, perhaps the majority, are probably autoimmune in origin. Cases in which hepatotropic viruses, metabolic/genetic derangements (including non-alcoholic fatty liver disease), and hepatotoxic drugs have been excluded represent a spectrum of heterogeneous liver disorders of unknown cause, a proportion of which are most likely autoimmune hepatitis.

The weight of evidence suggests that the progressive liver injury in patients with autoimmune hepatitis is the result of a cell-mediated immunologic attack directed against liver cells in the setting of a loss of, or failed, immunologic tolerance for self-liver antigens. In all likelihood, predisposition to autoimmunity is inherited, whereas the liver specificity of this injury is triggered by environmental (e.g., chemical, drug [e.g., minocycline], or viral) factors. For example, patients have been described in whom apparently self-limited cases of acute hepatitis A, B, or C led to autoimmune hepatitis, presumably because of genetic susceptibility or predisposition. Evidence to support an autoimmune pathogenesis in this type of hepatitis includes the following: (1) in the liver, the histopathologic lesions are composed predominantly of cytotoxic T cells and plasma cells; (2) circulating autoantibodies (nuclear, smooth muscle, thyroid, etc.; see below), rheumatoid factor, and hyperglobulinemia are common; (3) other autoimmune disorders—such as autoimmune thyroiditis, rheumatoid arthritis, autoimmune hemolytic anemia, ulcerative colitis, membranoproliferative glomerulonephritis, juvenile diabetes mellitus, vitiligo, celiac disease, and Sjögren's syndrome—occur with increased frequency in patients who have autoimmune hepatitis and in their relatives; (4) histocompatibility haplotypes associated with autoimmune diseases, such as HLA-B1, B8, DR3, and DR4 as well as extended haplotype *DRB1 0301* and *DRB1 0401* alleles, are common in patients with autoimmune hepatitis; and (5) this type of chronic hepatitis is responsive to glucocorticoid/immunosuppressive therapy, effective in a variety of autoimmune disorders.

Cellular immune mechanisms appear to be important in the pathogenesis of autoimmune hepatitis. In vitro studies have suggested that in patients with this disorder, CD4+ T lymphocytes are capable of becoming sensitized to hepatocyte membrane proteins and of destroying liver cells. Molecular mimicry by cross-reacting antigens that contain epitopes similar to liver antigens is postulated to activate these T cells, which infiltrate, and result in injury to, the liver. Abnormalities of immunoregulatory control over cytotoxic lymphocytes (impaired regulatory CD4+CD25+ T-cell influences) may play a role as well. Studies of genetic predisposition to autoimmune hepatitis demonstrate that certain haplotypes are associated with the disorder, as enumerated above, as are polymorphisms in cytotoxic T lymphocyte antigens (*CTLA-4*) and tumor necrosis factor α (*TNFA 2*). The precise triggering factors, genetic influences, and cytotoxic and immunoregulatory mechanisms involved in this type of liver injury remain incompletely defined.

Intriguing clues into the pathogenesis of autoimmune hepatitis come from the observation that circulating autoantibodies are prevalent in patients with this disorder. Among the autoantibodies described in these patients are antibodies to nuclei (so-called antinuclear antibodies [ANAs], primarily in a homogeneous pattern) and smooth muscle (so-called anti-smooth-muscle antibodies, directed at actin, vimentin, and skeleton), antibodies to F-actin, anti-LKM (see below), antibodies to "soluble liver antigen" (directed against a uracil-guanine-adenine transfer RNA suppressor protein), antibodies to α -actinin, and antibodies to the liver-specific asialoglycoprotein receptor (or "hepatocyte lectin") and other hepatocyte membrane proteins. Although some of these provide helpful diagnostic markers, their involvement in the pathogenesis of autoimmune hepatitis has not been established.

Humoral immune mechanisms have been shown to play a role in the extrahepatic manifestations of autoimmune and idiopathic hepatitis. Arthralgias, arthritis, cutaneous vasculitis, and glomerulonephritis occurring in patients with autoimmune hepatitis appear to be mediated by the deposition of circulating immune complexes in affected tissue vessels, followed by complement activation, inflammation, and tissue injury. While specific viral antigen-antibody complexes can be identified in acute and chronic viral hepatitis, the nature of the immune complexes in autoimmune hepatitis has not been defined.

■ CLINICAL FEATURES

Many of the clinical features of autoimmune hepatitis are similar to those described for chronic viral hepatitis. The onset of disease may be insidious or abrupt; the disease may present initially like, and be

confused with, acute viral hepatitis; a history of recurrent bouts of what had been labeled *acute hepatitis* is not uncommon. In approximately a quarter of patients, the diagnosis is made in the absence of symptoms, based on abnormal liver laboratory tests. A subset of patients with autoimmune hepatitis has distinct features. Such patients are predominantly young to middle-aged women with marked hyperglobulinemia and high titer circulating ANAs. This is the group with positive lupus erythematosus (LE) preparations (initially labeled *lupoid hepatitis*) in whom other autoimmune features are common. Fatigue, malaise, anorexia, amenorrhea, acne, arthralgias, and jaundice are common. Occasionally, arthritis, maculopapular eruptions (including cutaneous vasculitis), erythema nodosum, colitis, pleuritis, pericarditis, anemia, azotemia, and sicca syndrome (keratoconjunctivitis, xerostomia) occur. In some patients, complications of cirrhosis, such as ascites and edema (associated with portal hypertension and hypoalbuminemia), encephalopathy, hypersplenism, coagulopathy, or variceal bleeding may bring the patient to initial medical attention.

The course of autoimmune hepatitis may be variable. In patients with mild disease or limited histologic lesions (e.g., piecemeal necrosis [inflammation and erosion of the limiting place of periportal hepatocytes] without bridging), progression to cirrhosis is limited, but, even in this subset, clinical monitoring is important to identify progression; up to half left untreated can progress to cirrhosis over the course of 15 years. In North America, cirrhosis at presentation is more common in African Americans than in whites. In those with severe symptomatic autoimmune hepatitis (aminotransferase levels >10 times normal, marked hyperglobulinemia, "aggressive" histologic lesions—bridging necrosis or multilobular collapse, cirrhosis), the 6-month mortality without therapy may be as high as 40%. Such severe disease accounts for only 20% of cases; the natural history of milder disease is variable, often accentuated by spontaneous remissions and exacerbations. In a 10-year (2006–2016) national Dutch study, mortality in patients with autoimmune hepatitis was higher than that of the general population only in patients with cirrhosis; for patients without cirrhosis, survival was comparable to that of the general population. Especially poor prognostic signs include the presence histologically of multilobular collapse at the time of initial presentation and failure of serum bilirubin to improve after 2 weeks of therapy. Death may result from hepatic failure, hepatic coma, other complications of cirrhosis (e.g., variceal hemorrhage), and intercurrent infection. In patients with established cirrhosis, HCC may be a late complication (Chap. 82) but occurs less frequently than in cirrhosis associated with viral hepatitis.

Laboratory features of autoimmune hepatitis are similar to those seen in chronic viral hepatitis. Liver biochemical tests are invariably abnormal but may not correlate with the clinical severity or histopathologic features in individual cases. Many patients with autoimmune hepatitis have normal serum bilirubin, alkaline phosphatase, and globulin levels with only minimal aminotransferase elevations. Serum AST and ALT levels are increased and fluctuate in the range of 100–1000 units. In severe cases, the serum bilirubin level is moderately elevated (51–171 $\mu\text{mol/L}$ [3–10 mg/dL]). Hypoalbuminemia occurs in patients with very active or advanced disease. Serum alkaline phosphatase levels may be moderately elevated or near normal. In a small proportion of patients, marked elevations of alkaline phosphatase activity occur; in such patients, clinical and laboratory features overlap with those of primary biliary cholangitis (Chap. 344). The prothrombin time is often prolonged, particularly late in the disease or during active phases.

Polyclonal hypergammaglobulinemia (>2.5 g/dL) is common in autoimmune hepatitis, as is the presence of rheumatoid factor. As noted above, circulating autoantibodies are also prevalent, most characteristically ANAs in a homogeneous staining pattern. Smooth-muscle antibodies are less specific, seen just as frequently in chronic viral hepatitis. Because of the high levels of globulins achieved in the circulation of some patients with autoimmune hepatitis, occasionally the globulins may bind nonspecifically in solid-phase binding immunoassays for viral antibodies. This has been recognized most commonly in tests for antibodies to HCV, as noted above. In fact, studies of autoantibodies in autoimmune hepatitis have led to the recognition of new categories

of autoimmune hepatitis. *Type I autoimmune hepatitis* is the classic syndrome prevalent in North America and northern Europe occurring in young women, associated with marked hyperglobulinemia, lupoid features, circulating ANAs, and HLA-DR3 or HLA-DR4 (especially *B8-DRB1 03*). Also associated with type I autoimmune hepatitis are autoantibodies against actin and atypical perinuclear antineutrophilic cytoplasmic antibodies (pANCA). Included in the spectrum of type I autoimmune hepatitis is a subset of patients who lack ANA and anti-LKM1 but who have circulating antibodies to soluble liver antigen. Most of these patients are women and have clinical features similar to, or perhaps more severe than, those of other patients with type I autoimmune hepatitis.

Type II autoimmune hepatitis, often seen in children, more common in Mediterranean populations, and linked to HLA-DRB1 and HLA-DQB1 haplotypes, is associated not with ANA but with anti-LKM. Actually, anti-LKM represent a heterogeneous group of antibodies. In type II autoimmune hepatitis, the antibody is anti-LKM1, directed against cytochrome P450 2D6. This is the same anti-LKM seen in some patients with chronic hepatitis C. Anti-LKM2 is seen in drug-induced hepatitis, and anti-LKM3 (directed against uridine diphosphate glucuronyltransferases) is seen in patients with chronic hepatitis D. Another autoantibody observed in type II autoimmune hepatitis is directed against liver cytosol formiminotransferase cyclodeaminase (anti-liver cytosol 1).

Liver biopsy abnormalities are similar to those described for chronic viral hepatitis. Expanding portal tracts and extending beyond the plate of periportal hepatocytes into the parenchyma (designated *interface hepatitis* or *piecemeal necrosis*) is a mononuclear cell infiltrate that, in autoimmune hepatitis, may include the presence of plasma cells. Necroinflammatory activity characterizes the lobular parenchyma, and evidence of hepatocellular regeneration is reflected by "rosette" formation, the occurrence of thickened liver cell plates, and regenerative "pseudolobules." Septal fibrosis, bridging fibrosis, and cirrhosis are frequent. In patients with early autoimmune hepatitis presenting as an acute-hepatitis-like illness, lobular and centrilobular (as opposed to the more common periportal) necrosis has been reported. Bile duct injury and granulomas are uncommon; however, a subgroup of patients with autoimmune hepatitis has histologic, biochemical, and serologic features overlapping those of primary biliary cholangitis (Chap. 344).

■ DIAGNOSTIC CRITERIA

An international group has suggested a set of criteria for establishing a diagnosis of autoimmune hepatitis. Exclusion of liver disease caused by genetic disorders, viral hepatitis, drug hepatotoxicity, and alcohol is linked with such inclusive diagnostic criteria as hyperglobulinemia, autoantibodies, and characteristic histologic features. This international group has also suggested a comprehensive diagnostic scoring system that, rarely required for typical cases, may be helpful when typical features are not present. Factors that weigh in favor of the diagnosis include female gender; predominant aminotransferase elevation; presence and level of globulin elevation; presence of nuclear, smooth-muscle, LKM1, and other autoantibodies; concurrent other autoimmune diseases; characteristic histologic features (interface hepatitis, plasma cells, rosettes); HLA-DR3 or DR4 markers; and response to treatment (see below). A more simplified, more specific scoring system relies on four variables: autoantibodies, serum IgG level, typical or compatible histologic features, and absence of viral hepatitis markers. Weighing against the diagnosis are predominant alkaline phosphatase elevation, mitochondrial antibodies, markers of viral hepatitis, history of hepatotoxic drugs or excessive alcohol, histologic evidence of bile duct injury, or such atypical histologic features as fatty infiltration, iron overload, and viral inclusions.

■ DIFFERENTIAL DIAGNOSIS

Early during the course of chronic hepatitis, autoimmune hepatitis may resemble typical *acute viral hepatitis* (Chap. 339). Without histologic assessment, severe chronic hepatitis cannot be readily distinguished based on clinical or biochemical criteria from mild chronic hepatitis. In adolescence, *Wilson's disease* (Chaps. 344 and 415) may present

with features of chronic hepatitis long before neurologic manifestations become apparent and before the formation of *Kayser-Fleischer rings* (copper deposition in Descemet's membrane in the periphery of the cornea). In this age group, serum ceruloplasmin and serum and urinary copper determinations plus measurement of liver copper levels establish the correct diagnosis. *Postnecrotic* or *cryptogenic cirrhosis* and *primary biliary cholangitis* (Chap. 344) share clinical features with autoimmune hepatitis, and both alcoholic hepatitis (Chap. 342) and nonalcoholic steatohepatitis (Chap. 343) may present with many features common to autoimmune hepatitis; historic, biochemical, serologic, and histologic assessments are usually sufficient to allow these entities to be distinguished from autoimmune hepatitis. Of course, the distinction between autoimmune and chronic viral hepatitis is not always straightforward, especially when viral antibodies occur in patients with autoimmune disease or when autoantibodies occur in patients with viral disease. Furthermore, the presence of extrahepatic features such as arthritis, cutaneous vasculitis, or pleuritis—not to mention the presence of circulating autoantibodies—may cause confusion with *rheumatologic disorders* such as rheumatoid arthritis and systemic LE. The existence of clinical and biochemical features of progressive necroinflammatory liver disease distinguishes chronic hepatitis from these other disorders, which are not associated with severe liver disease. Rarely, hepatic venous outflow obstruction (Budd-Chiari syndrome) may present with features suggestive of autoimmune hepatitis, but painful hepatomegaly, ascites, and vascular imaging provide distinguishing diagnostic clues. Other diagnostic considerations would include celiac disease and ischemic liver disease, which would be readily distinguishable by clinical and laboratory features from autoimmune hepatitis.

In patients treated with immune checkpoint inhibitors for malignancy, the liver may be one of the autoimmune targets of therapy; the syndrome resembles autoimmune hepatitis in clinical features and response to glucocorticoid-based treatment. Finally, occasionally, features of autoimmune hepatitis overlap with features of autoimmune biliary disorders such as primary biliary cholangitis, primary sclerosing cholangitis (Chaps. 344 and 346), or, even more rarely, mitochondrial antibody-negative autoimmune cholangitis. Such overlap syndromes are difficult to categorize, and often response to therapy may be the distinguishing factor that establishes the diagnosis.

TREATMENT

Autoimmune Hepatitis

The mainstay of management in autoimmune hepatitis is glucocorticoid therapy. Several controlled clinical trials have documented that such therapy leads to symptomatic, clinical, biochemical, and histologic improvement as well as increased survival. A therapeutic response can be expected in up to 80% of patients. Unfortunately, therapy has not been shown in clinical trials to prevent ultimate progression to cirrhosis; however, instances of reversal of fibrosis and cirrhosis have been reported in patients responding to treatment, and rapid treatment responses within 1 year do translate into a reduction in progression to cirrhosis. Although some advocate the use of prednisolone (the hepatic metabolite of prednisone), prednisone is just as effective and is favored by most authorities. Therapy may be initiated at 20 mg/d, but a popular regimen in the United States relies on an initiation dose of 60 mg/d. This high dose is tapered successively over the course of a month down to a maintenance level of 20 mg/d. An alternative, but equally effective, more appealing approach is to begin with half the prednisone dose (30 mg/d) along with azathioprine (50 mg/d). With azathioprine maintained at 50 mg/d, the prednisone dose is tapered over the course of a month down to a maintenance level of 10 mg/d. The advantage of the combination approach is a reduction, over the span of an 18-month course of therapy, in serious, life-threatening complications of steroid therapy (e.g., cushingoid features, hypertension, diabetes, osteoporosis) from 66% down to <20%. Genetic analysis for thiopurine S-methyltransferase allelic variants does not

correlate with azathioprine-associated cytopenias or efficacy and is not assessed routinely in patients with autoimmune hepatitis. In combination regimens, 6-mercaptopurine may be substituted for its prodrug azathioprine, but this is rarely required. Azathioprine alone, however, is not effective in achieving remission, nor is alternate-day glucocorticoid therapy. Limited experience with budesonide in noncirrhotic patients suggests that this steroid side effect–sparing drug may be effective; however, the few randomized controlled trials of budesonide have not consistently shown efficacy. Although therapy has been shown to be effective for severe autoimmune hepatitis (AST $\geq 10 \times$ the upper limit of normal or $\geq 5 \times$ the upper limit of normal in conjunction with serum globulin greater than or equal to twice normal; bridging necrosis or multi-lobular necrosis on liver biopsy; presence of symptoms), therapy is not indicated for mild forms of chronic hepatitis, and the efficacy of therapy in mild or asymptomatic autoimmune hepatitis has not been established.

Improvement of fatigue, anorexia, malaise, and jaundice tends to occur within days to several weeks; biochemical improvement occurs over the course of several weeks to months, with a fall in serum bilirubin and globulin levels and an increase in serum albumin. Serum aminotransferase levels usually drop promptly, but improvements in AST and ALT alone do not appear to be reliable markers of recovery in individual patients; histologic improvement, characterized by a decrease in mononuclear infiltration and in hepatocellular necrosis, may be delayed for 6–24 months. Still, if interpreted cautiously, aminotransferase levels are valuable indicators of relative disease activity, and, although recommended, many authorities do *not* advocate for serial liver biopsies to assess therapeutic success or to guide decisions to alter or stop therapy. Rapidity of response is more common in older patients (≥ 69 years) and those with HLA *DBR1 04*; although rapid responders may progress less slowly to cirrhosis and liver transplantation, they are no less likely than slower responders to relapse after therapy. Therapy should continue for at least 12–18 months. After tapering and cessation of therapy, the likelihood of relapse is at least 50%, even if posttreatment histology has improved to show mild chronic hepatitis, and most patients require therapy at maintenance doses indefinitely. Continuing azathioprine alone (2 mg/kg body weight daily) after cessation of prednisone therapy has been shown to reduce the frequency of relapse. Long-term maintenance with low-dose prednisone (≤ 10 mg daily) has also been shown to keep autoimmune hepatitis in check without the theoretical risk of azathioprine marrow suppression and, in young women of child-bearing age, teratogenicity; however, maintenance azathioprine is more effective in preserving remission.

In medically refractory cases, an attempt should be made to intensify treatment with high-dose glucocorticoid monotherapy (60 mg daily) or combination glucocorticoid (30 mg daily) plus high-dose azathioprine (150 mg daily) therapy. After a month, doses of prednisone can be reduced by 10 mg a month, and doses of azathioprine can be reduced by 50 mg a month toward ultimate, conventional maintenance doses. Patients refractory to this regimen may be treated with cyclosporine, tacrolimus, or mycophenolate mofetil. Similarly, in exploratory studies, infusions of monoclonal antibodies directed at tumor necrosis factor (infliximab) and against the B-lymphocyte antigen CD20 (rituximab) have been reported to be of clinical benefit (improved aminotransferase levels, immunoglobulin G levels, histologic inflammatory activity) as rescue therapy for refractory autoimmune hepatitis. To date, however, only limited, often anecdotal, data in small numbers of patients support these alternative approaches. If medical therapy fails, or when chronic hepatitis progresses to cirrhosis and is associated with life-threatening complications of liver decompensation, liver transplantation is the only recourse (Chap. 345); in patients with severe autoimmune hepatitis, failure of the bilirubin to improve after 2 weeks of therapy should prompt early consideration of the patient for liver transplantation. Recurrence of autoimmune hepatitis in the new liver occurs rarely in most experiences but in as many

as 35–40% of cases in others; nonetheless, 5-year patient and graft survival exceed 80%.

Like all patients with chronic liver disease, patients with autoimmune hepatitis should be vaccinated against hepatitis A and B, ideally before immunosuppressive therapy is begun, if practical. Patients with autoimmune hepatitis and cirrhosis should be screened for HCC with ultrasound at 6-month intervals and for gastroesophageal varices with upper gastrointestinal endoscopy at intervals of 1–3 years, based on severity of liver disease.

FURTHER READING

- AASLD/IDSA HCV G : Hepatitis C guidance 2019 update: American Association for the Study of Liver Diseases–Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. *Hepatology* 71:686, 2020. Updated regularly and available at <http://www.hcvguidelines.org>. Accessed April 20, 2020.
- B -M SI et al: Hepatitis B virus reactivation associated with direct-acting antiviral therapy for chronic hepatitis C virus: A review of cases reported to the U.S. Food and Drug Administration Adverse Event Reporting System. *Ann Intern Med* 166:792, 2017.
- B M et al: Sofosbuvir, velpatasvir, and voxilaprevir for previously treated HCV infection. *N Engl J Med* 376:2134, 2017.
- B M et al: Tenofovir alafenamide versus tenofovir disoproxil fumarate for the treatment of HBeAg-negative chronic hepatitis B virus infection: A randomized, double-blind, phase 3 non-inferiority trial. *Lancet Gastroenterol Hepatol* 1:196, 2017.
- B AA et al: Direct-acting antiviral therapy for HCV infection is associated with a reduced risk of cardiovascular disease events. *Gastroenterology* 156:987, 2019.
- C M, N JM: Autoimmune liver disease, autoimmunity and liver transplantation. *J Hepatol* 60:210, 2014.
- C F et al: Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatments: A prospective cohort study. *Lancet* 393:1453, 2019.
- C HLY et al: Tenofovir alafenamide versus tenofovir disoproxil fumarate for the treatment of HBeAg-positive chronic hepatitis B virus infection: A randomized, double-blind, phase 3 non-inferiority trial. *Lancet Gastroenterol Hepatol* 1:185, 2017.
- E A S L : EASL 2017 clinical practice guidelines on the management of hepatitis B virus infection. *J Hepatol* 67:370, 2017.
- E A S L : EASL recommendations on treatment of hepatitis C 2018. *J Hepatol* 69:461, 2018.
- E A S L : EASL recommendations on treatment of hepatitis C: Final update of the series. *J Hepatol* 73:1170, 2020.
- F X et al: Glecaprevir plus pibrentasvir for chronic hepatitis C virus genotype 1, 2, 4, 5, or 6 infection in adults with compensated cirrhosis (EXPEDITION-1): A single-arm, open-label, multicentre phase 3 trial. *Lancet Infect Dis* 17:1062, 2017.
- J IM et al: American Gastroenterological Association Institute clinical practice update-expert review: Care of patients who have achieved a sustained virologic response after antiviral therapy for chronic hepatitis C infection. *Gastroenterology* 152:1578, 2017.
- K PY et al: Glecaprevir and pibrentasvir yield high response rates in patients with HCV genotype 1–6 without cirrhosis. *J Hepatol* 67:263, 2017.
- L KS et al: Limited sustained response after stopping nucleos(t)ide analogues in patients with chronic hepatitis B: Results from a randomized controlled trial (Toronto STOP study). *Gut* 68:2206, 2019.
- L ASG et al: Antiviral therapy for chronic hepatitis B viral infection in adults: A systematic review and meta-analysis. *Hepatology* 63:284, 2016.
- L R, L TJ: Hepatitis B reactivation associated with immune suppressive and biological modifier therapies: Current concepts, management strategies, and future directions. *Gastroenterology* 152:1297, 2017.

- M EA et al: Assessing the safety of direct-acting antiviral agents for hepatitis C. *JAMA Open Network* 2(6):e194765, 2019.
- P GV et al: Eight-year survival in chronic hepatitis B patients under long-term entecavir or tenofovir is similar to the general population. *J Hepatol* 68:1129, 2018.
- P GV et al: DARING-B: Discontinuation of effective entecavir or tenofovir disoproxil fumarate long-term therapy before HBsAg loss in non-cirrhotic HBeAg-negative chronic hepatitis B. *Antivir Ther* 23:677, 2018.
- P J-M et al: From non-A, non-B hepatitis to hepatitis C virus cure. *J Hepatol* 62:S87, 2015.
- P RP et al: American Gastroenterological Association Institute technical review on prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. *Gastroenterology* 148:221, 2015.
- R KJ et al: American Gastroenterological Association Institute guideline on the prevention and treatment of hepatitis B virus reactivation during immunosuppressive drug therapy. *Gastroenterology* 148:215, 2015.
- R C et al: Sustained virologic response from interferon-based hepatitis C regimens is associated with reduced risk of extrahepatic manifestations. *J Hepatol* 71:1116, 2019.
- S AG et al: AGA clinical practice update on interaction between oral direct-acting antivirals for chronic hepatitis C infection and hepatocellular carcinoma: Expert review. *Gastroenterology* 156:2149, 2019.
- S Set al: Magnitude and kinetics of decrease in liver stiffness after antiviral therapy in patients with chronic hepatitis C: A systematic review and meta-analysis. *Clin Gastroenterol Hepatol* 16:27, 2018.
- S CW et al: Hepatitis C. *Lancet* 394:1451, 2019.
- T LS et al: Chronic hepatitis B infection: A review. *JAMA* 319:1802, 2018.
- T N et al: Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* 67:1560, 2018.
- V D B FF et al: Increased mortality among patients with vs without cirrhosis and autoimmune hepatitis. *Clin Gastroenterol Hepatol* 17:940, 2019.
- Y C et al: Treating chronic hepatitis delta: The need for surrogate markers of treatment efficacy. *J Hepatol* 70:1008, 2019.

Liver cirrhosis is the eleventh leading cause of mortality worldwide, causing 1.16 million deaths annually; 48% of cases of cirrhosis can be attributed to alcohol. Among patients with alcohol use disorder, 18% had fibrosis, 26% had cirrhosis, and 7% had acute alcoholic hepatitis without underlying cirrhosis. In the European population, the annual incidence rate for acute alcoholic hepatitis is between 24 and 27 per million persons in women and between 46 and 65 per million persons in men.

■ PATHOGENESIS

Alcohol in the form of ethanol is rapidly absorbed in the upper gastrointestinal tract and predominantly metabolized in the liver. Ethanol reaches the liver through the portal vein, and the majority of ethanol is oxidized via alcohol dehydrogenase 1 (ADH1) into acetaldehyde in hepatocytes. Chronic alcohol consumption induces the expression of a second ethanol-metabolizing enzyme, cytochrome P450 family 2 subfamily E member 1 (CYP2E1), which also converts ethanol into acetaldehyde. In addition to the direct cellular toxic effects of acetaldehyde, metabolism of ethanol into acetaldehyde causes the generation of reactive oxygen species (ROS), resulting in further injury of hepatocytes via lipid peroxidation and DNA damage. Acetaldehyde is then oxidized into acetate via acetaldehyde dehydrogenase (ALDH). Inherited deficiency of ALDH2 is common in Asian countries and leads to acetaldehyde accumulation after alcohol consumption. These individuals develop nausea and cutaneous flushing. Several mechanisms contribute to the development of hepatic steatosis related to alcohol consumption. Acetate is converted into acetyl-coenzyme A (CoA), which contributes to fatty acid and triglyceride synthesis. Alcohol, in part through epigenetic changes, increases the expression of genes involved in lipogenesis, while genes involved in fatty acid transport and oxidation are suppressed. Alcohol also increases the ratio of reduced nicotinamide adenine dinucleotide (NAD)/oxidized NAD (NADH/NAD⁺) in hepatocytes, which further reduces mitochondrial β-oxidation. Alcohol can increase fatty acid mobilization in adipose tissue and the intestine, which will lead to hepatic accumulation of fatty acids and increased hepatic steatosis. Overall, the net effect of these processes contributes to fat accumulation in the liver.

■ RISK FACTORS FOR PROGRESSION OF ALD

Daily alcohol consumption or heavy drinking results in hepatic steatosis, but only 10–20% of such individuals will develop progressive liver disease and cirrhosis. Therefore, other cofactors such as behavioral, environmental, and genetic factors play important roles in progression of ALD (Table 342-1). There is a dose-dependent increase, with regard to the amount of alcohol consumed, in the likelihood of developing liver cirrhosis. Women develop ALD at a lower daily alcohol intake. Cigarette smoking is an independent risk factor for alcohol-associated cirrhosis. The drinking pattern, in particular binge drinking and excessive alcohol drinking outside meals, increases the risk of developing progressive ALD. Obesity and other chronic liver diseases such as viral hepatitis, hemochromatosis, and nonalcoholic steatohepatitis (NASH), are frequent cofactors contributing to progression of ALD. Twin studies demonstrated a genetic predisposition to alcohol-associated liver cirrhosis that is independent from the genetic predisposition to alcohol use disorder. Gene polymorphisms conferring increased risk of alcohol-associated liver cirrhosis have been found in three genes, patatin-like phospholipase domain-containing 3 (*PNPLA3*),

342 Alcohol-Associated Liver Disease

Bernd Schnabl



Alcohol-associated liver diseases (ALD) comprise a spectrum of diseases associated with chronic alcohol consumption ranging from alcohol-associated fatty liver disease and steatohepatitis to more advanced liver disease including fibrosis and cirrhosis. Acute alcoholic hepatitis is an acute-on-chronic form of ALD that is associated with liver failure and high mortality.

■ EPIDEMIOLOGY

Approximately 5.8% of adults in the United States have an alcohol use disorder, defined as >2 drinks per day in women and >3 drinks per day in men, or partake in binge drinking, defined as 4 drinks for women and 5 drinks for men in ~2 h (1 drink equals ~14 g of ethanol, which is 1 beer, 4 oz of wine, or 1 oz of 80% spirits). Prevalence of ALD correlates with the amount of alcohol consumption in different regions. Prevalence of alcohol-associated fatty liver disease is 4.7% of the general population in the United States, and 1.5% has stage 2 or greater fibrosis.

TABLE 342-1 Factors for Progression of Alcohol-Associated Liver Disease

- Alcohol dose (>1 drink per day for women, >2 drinks per day for men)
- Drinking pattern (drinking without meal, binge drinking)
- Genetic factors, especially *PNPLA3* polymorphism
- Female gender
- Smoking
- Increased body mass index and chronic liver diseases
- Intestinal microbiota

TABLE 342-2 Symptoms and Signs Associated with Alcohol-Associated Cirrhosis and Alcoholic Hepatitis

- Tiredness
- Malnutrition and sarcopenia
- Abdomen: abdominal discomfort, hepatomegaly, splenomegaly, caput medusae, ascites with weight gain, abdominal pain, and shortness of breath
- Skin: spider angioma, palmar erythema, jaundice, ecchymoses
- Eyes: icteric sclerae
- Hands: Dupuytren contracture
- Face: rhinophyma
- Reproductive system: gynecomastia, gonadal atrophy, loss of libido, amenorrhea
- Neurologic:
 - Peripheral neuropathy
 - Alcohol withdrawal: tachycardia, agitation, tremor, seizures, delirium
 - Hepatic encephalopathy: asterixis (flapping tremor), forgetfulness, inversion of sleep/wake pattern, altered consciousness, confusion, lethargy, coma
 - Wernicke-Korsakoff syndrome

membrane bound O-acyltransferase domain-containing 7 (*MBOAT7*), and transmembrane 6 superfamily member 2 (*TM6SF2*), although the molecular mechanism is not well understood. A subset of patients with alcohol use disorder develop changes in the gut microbiome and increased intestinal permeability resulting in activation of hepatic inflammation, hepatocyte death, and activation of fibrotic pathways. Ongoing fibrosis due to continued alcohol consumption can result in the development of cirrhosis with portal hypertension (Chap. 344).

■ CLINICAL FEATURES

The development of alcohol-associated steatosis, steatohepatitis, and cirrhosis is most often clinically silent. Symptoms arise once the patient with alcohol-associated liver cirrhosis decompensates or develops alcoholic hepatitis (Table 342-2). Patients with alcoholic hepatitis have been drinking heavily for typically >5 years and until at least 8 weeks before onset of symptoms. They present with rapid onset of jaundice (serum bilirubin >3 mg/dL), often accompanied by fever, malaise, tender hepatomegaly, and clinical signs of hepatic decompensation, such as ascites, bacterial infection, variceal bleeding, and hepatic encephalopathy. Infections occur in 12–26% of patients with severe alcoholic hepatitis at the time of admission. Alcoholic hepatitis is often accompanied by systemic inflammatory response syndrome (SIRS) and acute kidney injury (AKI) secondary to hepatorenal syndrome.

■ LABORATORY FINDINGS

Patients with simple hepatic steatosis can present with normal liver function tests. Steatohepatitis is characterized by elevated levels of aspartate aminotransferase (AST) and γ -glutamyl transferase (GGT). Characteristic laboratory parameters for ALD include a ratio of AST to alanine aminotransferase (ALT) of >1, and serum AST is rarely >300 IU/L. Serum bilirubin and international normalized ratio (INR) are typically normal. Elevated bilirubin and INR and low serum albumin and platelet count are common laboratory findings in patients with cirrhosis. Patients with alcoholic hepatitis have AST and ALT elevations that do not exceed 400 IU/L, with AST/ALT ratio of >1.5 and serum bilirubin >3 mg/dL.

■ DIAGNOSIS

The Alcohol Use Disorders Inventory Test (AUDIT) is a validated tool for identifying patients with alcohol use disorder (Chap. 453). Diagnosis of ALD requires exclusion of other liver diseases in heavy drinkers. Alcohol-associated steatosis can be diagnosed by simple ultrasound, magnetic resonance imaging (MRI), or computed tomography (CT). Noninvasive quantification of hepatic fat can be achieved with the ultrasound technique of controlled attenuation parameter (CAP) or with magnetic resonance proton density fat fraction (MR-PDFF). Liver biopsy is rarely indicated for diagnosing alcohol-associated hepatic steatosis or steatohepatitis. Liver biopsy typically shows hepatocytes

with large lipid droplets (macrovesicular steatosis) around pericentral veins (zone 3). Morphologic features of alcohol-associated steatohepatitis include hepatocyte injury and ballooning with Mallory-Denk bodies, necrosis, and lobular inflammation with mononuclear and neutrophilic granulocytes.

Progression of alcohol-associated steatohepatitis to fibrosis can be diagnosed using liver stiffness measurement by techniques such as transient elastography (e.g., FibroScan). Liver stiffness <6 kPa indicates normal liver, whereas cutoffs for each stage of alcohol-associated liver fibrosis have been validated (>8 kPa indicates F3 advanced fibrosis; >12.5 kPa indicates F4 cirrhosis). Histology shows initially perivenular fibrosis with subsequent extension of collagen fibers into hepatic lobules, described as septal fibrosis. Patients with cirrhosis show liver nodularity on imaging with ultrasound, MRI, or CT scan. Radiologic signs of portal hypertension include ascites, splenomegaly, and portal-systemic collateral vessels. Prognosis and risk of mortality are assessed using Child-Pugh-Turcotte (CPT) or Model for End-Stage Liver Disease (MELD; or sodium-MELD) scores (Chap. 344).

In patients presenting with features suggestive of alcoholic hepatitis, imaging is obtained to exclude biliary obstruction and hepatocellular carcinoma (HCC). In addition, other causes of liver disease such as viral hepatitis, Wilson's disease, and severe autoimmune liver disease should be ruled out. Histology shows macrovesicular steatosis, hepatocyte ballooning with Mallory-Denk bodies, megamitochondria, neutrophil infiltration, bilirubinostasis, and chicken wire fibrosis. The majority of patients with alcoholic hepatitis have underlying cirrhosis (80%) (Chap. 344), and 10–20% of patients with a clinical diagnosis of alcoholic hepatitis will have other liver diseases on biopsy. Therefore, in the presence of potential confounding factors, including possible ischemic hepatitis (in the setting of, e.g., hypotension, massive gastrointestinal bleeding, recent cocaine use, septic shock), drug-induced liver injury (DILI), autoimmune liver disease, uncertain alcohol use assessment, or atypical laboratory tests (AST <50 IU/L or >400 IU/L, AST/ALT ratio <1.5), a transjugular liver biopsy is recommended to confirm the diagnosis of alcoholic hepatitis. Infections need to be assessed routinely with chest x-ray and blood, urine, and ascites cultures in patients presenting with alcoholic hepatitis.

TREATMENT

Alcohol-Associated Liver Disease (Fig. 342-1)

To date, the most effective therapy to reduce the progression of and reverse ALD is prolonged alcohol abstinence. In particular, alcohol-associated hepatic steatosis and steatohepatitis are reversible with cessation of alcohol consumption. Thus, treatment of the underlying alcohol use disorder is an integral part for therapy of ALD. There are currently no approved drugs for treatment of alcohol-associated steatosis and steatohepatitis with or without fibrosis.

Patients with alcohol-associated cirrhosis and ongoing alcohol consumption are at risk for decompensation and development of hepatic encephalopathy, ascites, variceal bleeding, hepatorenal syndrome, and HCC (Chap. 344). Patients with cirrhosis should undergo an upper gastrointestinal endoscopy to screen for varices. HCC screening is recommended using ultrasonography every 6 months in patients with cirrhosis. Management of complications of cirrhosis such as variceal bleeding, ascites, hepatic encephalopathy, and HCC does not differ from patients with cirrhosis due to a different etiology (Chap. 344). Liver transplantation for patients with alcohol-associated decompensated cirrhosis or HCC is a definitive therapy and is currently the leading indication for liver transplantation in the United States. Liver transplantation evaluation should be taken into consideration for patients with end-stage liver disease (Chap. 345).

In patients diagnosed with alcoholic hepatitis, short-term mortality can be predicted using the Maddrey discriminant function (MDF; calculated as $4.6 \times [\text{the prolongation of the prothrombin time above control [seconds]}] + \text{serum bilirubin [mg/dL]}$), MELD score (Chap. 344), or age-bilirubin-INR-creatinine (ABIC) score.

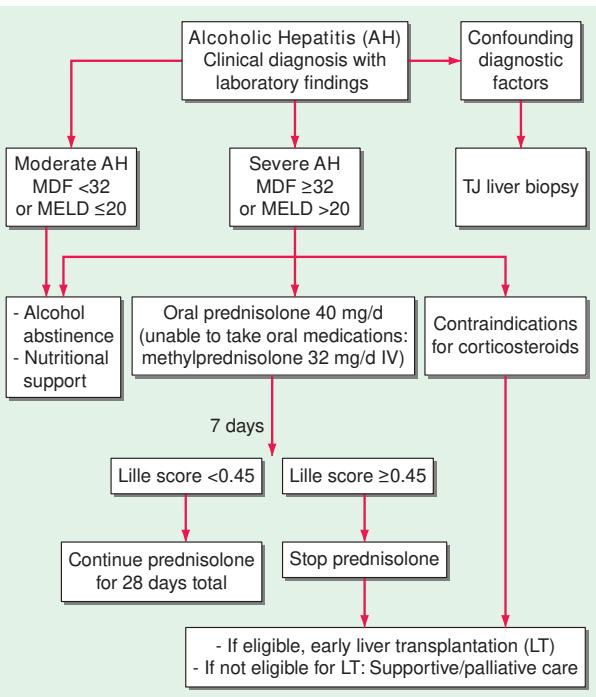


FIGURE 342-1 Treatment algorithm for alcoholic hepatitis. In patients with a clinical diagnosis of alcoholic hepatitis, confounding factors (see text) need to be ruled out, if necessary, by transjugular (TJ) liver biopsy. Patients with severe alcoholic hepatitis (AH), defined as Maddrey discriminant function (MDF) ≥ 32 or Model for End-Stage Liver Disease (MELD) score > 20 , without contraindications for glucocorticoids (see text) are candidates for such treatment. Nonresponders or patients with contraindications for treatment should be considered for early liver transplantation (LT) or supportive or palliative care, as clinically appropriate.

Patients with MDF < 32 or MELD ≤ 20 are defined as having moderate alcoholic hepatitis. Currently, patients with moderate alcoholic hepatitis are treated under a multidisciplinary team including an alcohol use disorder specialist, dietitian for nutritional supplementation for patients with markedly reduced intake, and hepatologist for managing liver disease complications. Enteral nutrition with a goal of > 21 kcal/kg and supplementation of micronutrients (in particular zinc) and vitamin supplementation (in particular vitamin B₁) are recommended for patients with alcoholic hepatitis. Intravenous albumin is preferred for volume expansion. MDF ≥ 32 or MELD > 20 identifies patients with severe alcoholic hepatitis and high short-term mortality who will have a survival benefit with glucocorticoid treatment. Contraindications for glucocorticoid treatment include uncontrolled infections or sepsis, AKI and hepatorenal syndrome, uncontrolled upper gastrointestinal bleeding, concomitant diseases (including viral hepatitis, HCC, pancreatitis, DILI, active tuberculosis, and HIV), multiorgan failure, and shock. Glucocorticoids can be used once infection, sepsis, and gastrointestinal bleeding are adequately controlled. Glucocorticoid use reduces the risk of death in patients with severe alcoholic hepatitis within 28 days of treatment but not in the following 6 months. Oral prednisolone, 40 mg/d for a total duration of 4 weeks, is preferred. For patients unable to take oral medications, methylprednisolone, 32 mg/d IV, is used. The combination of glucocorticoids with *N*-acetylcysteine infusion might add short-term survival benefit at 1 month. Failure of improvement of Lille score (≥ 0.45) after 7 days of glucocorticoid treatment will determine patients with severe alcoholic hepatitis who will unlikely benefit from continued treatment with glucocorticoids. Glucocorticoids should be stopped in nonresponders, and early liver transplantation should be considered. Although short-term prognosis is dependent on liver disease severity at the time of presentation, long-term prognosis (> 1 year) largely depends on

alcohol abstinence and underlying cirrhosis. Patients with severe alcoholic hepatitis that is nonresponsive to medical therapy have high 30-day mortality and are therefore unable to fulfill a minimum of 6 months of alcohol abstinence, which is required in many centers for liver transplantation evaluation. Early liver transplantation can be successfully performed in highly selected patients with an excellent psychosocial profile (Chap. 345). If a nonresponder is ineligible for early liver transplantation, supportive or palliative care should be considered for patients with multiple-organ failure.

FURTHER READING

- C DW et al: Standard definitions and common data elements for clinical trials in patients with alcoholic hepatitis: Recommendation from the NIAAA Alcoholic Hepatitis Consortia. Gastroenterology 150:785, 2016.
- C DW et al: Diagnosis and treatment of alcohol-associated liver diseases: 2019 practice guidance from the American Association for the Study of Liver Diseases. Hepatology 71:306, 2020.
- L A et al: Corticosteroids reduce risk of death within 28 days for patients with severe alcoholic hepatitis, compared with pentoxifylline or placebo—a meta-analysis of individual data from controlled trials. Gastroenterology 155:458, 2018.
- S HK et al: Alcoholic liver disease. Nat Rev Dis Primers 4:16, 2018.
- S AK et al: ACG clinical guideline: Alcoholic liver disease. Am J Gastroenterol 113:175, 2018.

343

Nonalcoholic Fatty Liver Diseases and Nonalcoholic Steatohepatitis

Manal F. Abdelmalek, Anna Mae Diehl

INCIDENCE, PREVALENCE, AND NATURAL HISTORY

Nonalcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease in the United States, as well as worldwide. The global prevalence of NAFLD is estimated to be as high as one billion. In the United States, NAFLD is estimated to affect between 80 and 100 million individuals. NAFLD is strongly associated with insulin resistance, overweight/obesity, and metabolic syndrome. However, it can also occur in lean individuals and is particularly common in those with a paucity of adipose depots (i.e., lipodystrophy). Ethnic/racial factors also appear to influence liver fat accumulation; the documented prevalence of NAFLD is lowest in African Americans (~25%), highest in Americans of Hispanic ancestry (~50%), and intermediate in American whites (~33%).

NAFLD encompasses a spectrum of liver pathology with different clinical prognoses (Fig. 343-1). The simple accumulation of triglyceride within hepatocytes (hepatocellular steatosis) is on the most clinically benign extreme of the spectrum. On the opposite, most clinically ominous extreme, are cirrhosis (Chap. 344) and primary liver cancer (Chap. 82). The risk of developing cirrhosis is extremely low in individuals with isolated steatosis (nonalcoholic fatty liver [NAFL]) but increases as steatosis becomes complicated by liver-cell injury and death and the accumulation of inflammatory cells (i.e., nonalcoholic steatohepatitis [NASH]). At least a quarter of adults with NAFLD are presumed to have NASH. NASH itself is also a heterogeneous condition; it can improve to steatosis or normal histology, remain relatively stable for years, or cause progressive accumulation of fibrous scar that eventuates in cirrhosis (stage 4 fibrosis). Advanced hepatic fibrosis is the primary predictor of eventual liver-related morbidity and mortality.

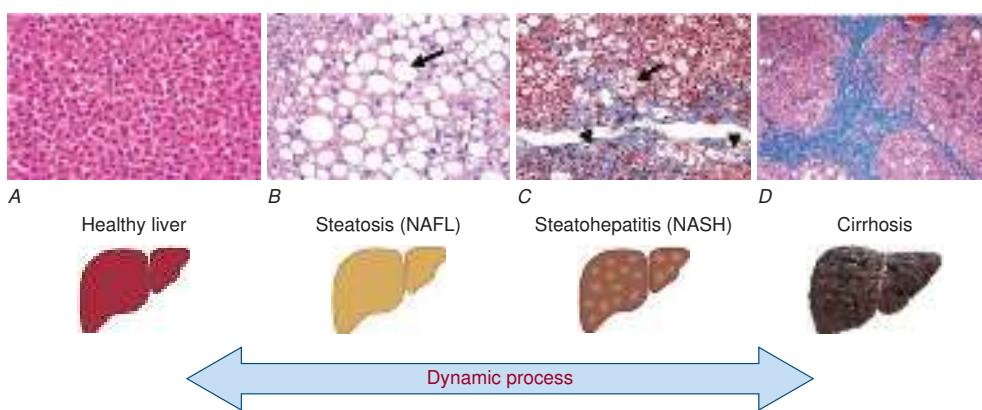


FIGURE 343-1 Histopathologic spectrum of nonalcoholic fatty liver disease (NAFLD). NAFLD encompasses a dynamic spectrum of liver pathology. **A.** Healthy liver. **B.** Simple steatosis (nonalcoholic fatty liver [NAFL]); arrow shows fatty hepatocyte. **C.** Nonalcoholic steatohepatitis (NASH); ballooned hepatocyte (arrow) near central vein with adjacent blue-stained pericellular fibrosis (arrowheads). **D.** Cirrhosis with blue-stained bridging fibrosis surrounding micronodules of liver parenchyma.

in NAFLD. Once NAFLD-related cirrhosis develops, the annual incidence of primary liver cancer can be as high as 1–2% per year.

Abdominal imaging is not able to determine which individuals with NAFLD have associated liver-cell death and inflammation (i.e., NASH), and specific blood tests to diagnose NASH are not yet available. However, population-based studies that have used elevated serum alanine aminotransferase (ALT) as a marker of liver injury indicate that ~6–8% of American adults have serum ALT elevations that cannot be explained by excessive alcohol consumption, other known causes of fatty liver disease (Table 343-1), viral hepatitis, or drug-induced or congenital liver diseases. Because the prevalence of such “cryptogenic”

ALT elevations increases with body mass index, it is presumed that they are due to NASH. Hence, at any given point in time, NASH is present in ~25% of individuals who have NAFLD (i.e., ~6–8% of the general U.S. adult population has NASH). Smaller cross-sectional studies in which liver biopsies have been performed on NASH patients at tertiary referral centers consistently demonstrate advanced fibrosis or cirrhosis in ~25% of those cohorts. By extrapolation, therefore,

TABLE 343-1 Alternative Causes of Hepatic Steatosis

- Alcoholic liver disease
- Hepatitis C (particularly genotype 3)
- Inborn errors of metabolism
 - Abetalipoproteinemia
 - Cholesterol ester storage disease
 - Galactosemia
 - Glycogen storage disease
 - Hereditary fructose intolerance
 - Homocystinuria
 - Systemic carnitine deficiency
 - Tyrosinemia
 - Weber-Christian syndrome
 - Wilson's disease
 - Wolman's disease
- Medications (see Table 343-2)
- Miscellaneous
 - Industrial exposure to petrochemical
 - Inflammatory bowel disease
 - Lipodystrophy
 - Bacterial overgrowth
 - Starvation
 - Parenteral nutrition
- Surgical procedures
 - Bilopancreatic diversion
 - Extensive small-bowel resection
 - Gastric bypass
 - Jejunoileal bypass
- Reye's syndrome
- Acute fatty liver of pregnancy
- HELLP syndrome (hemolytic anemia, elevated liver enzymes, low platelet count)

TABLE 343-2 Medications Associated with Hepatic Steatosis

- Cytotoxic and cytostatic drugs
 - 5-Fluorouraci
 - L-Asparaginase
 - Azacitidine
 - Azaserine
 - Bleomycin
 - Methotrexate
 - Puromycin
 - Tetracycline
 - Doxycycline
- Metals
 - Antimony
 - Barium salts
 - Chromates
 - Phosphorus
 - Rare earths of low atomic number
 - Thallium compounds
 - Uranium compounds
- Other drugs and toxins
 - Amiodarone
 - 4,4'-Diethylaminoethoxyhexesterol
 - Ethionine
 - Ethyl bromide
 - Estrogens
 - Glucocorticoids
- Highly active antiretroviral therapy
 - Hydralazine
 - Hypoglycin
 - Crotate
 - Perhexiline maleate
 - Safrole
 - Tamoxifen
 - Valproic acid
 - Acetylsalicylic acid intoxication
 - Apo-B inhibitors: Mipomersen and lomitapide

cirrhosis develops in ~6% of individuals with NAFLD (i.e., in ~1.5–2% of the general U.S. population). The risk for advanced liver fibrosis is highest in individuals with NASH who are aged >45–50 years and overweight/obese or afflicted with type 2 diabetes. Having a first-degree relative with cryptogenic hepatitis or cirrhosis also increases the risk for developing cirrhosis.

Heritable factors clearly impact susceptibility to hepatic steatosis, NASH, liver fibrosis, and liver cancer. Genetic variants on or near *TM6SF2* or *MBOAT7* (genes involved in lipid homeostasis) and palatin-like phospholipase domain-containing 3 gene (*PNPLA3*, a gene that encodes an enzyme involved in intracellular trafficking of lipids) may increase the heritability of NAFLD. A recent meta-analysis showed that *PNPLA3* exerts a strong influence not only on hepatic fat accumulation but also on the severity of NASH and liver fibrosis. Indeed, recent twin studies suggest that inheritance accounts for about half the risk for developing cirrhosis. Epigenetic factors (i.e., heritable traits that do not result from direct changes in DNA) may also influence NAFLD pathogenesis and/or progression based on evidence that intrauterine exposures influence susceptibility to obesity and the metabolic syndrome in adolescence. Studies of families with adult-onset obesity have identified genome-wide epigenetic alterations that dysregulate metabolic pathways controlling adiposity, insulin sensitivity, and tissue generation or regeneration. Whether such epigenetic mechanisms influence susceptibility to NASH and cirrhosis is being investigated.

NAFLD is currently the leading indication for liver transplantation in the United States. Similar to cirrhosis caused by other liver diseases, cirrhosis caused by NAFLD increases the risk for primary liver cancer. Both hepatocellular carcinoma and intrahepatic cholangiocarcinoma (ICC) have also been reported to occur in NAFLD patients without cirrhosis, suggesting that NAFLD per se may be a premalignant condition. NAFLD, NASH, and NAFLD-related cirrhosis are not limited to adults. All have been well documented in children. As in adults, obesity and insulin resistance are the main risk factors for pediatric NAFLD. Thus, the rising incidence and prevalence of childhood obesity suggests that NAFLD will be a major contributor to society's burden of liver disease in the future.

■ PATHOGENESIS

The mechanisms underlying the pathogenesis and progression of NAFLD are not entirely clear. The best-understood mechanisms pertain to hepatic steatosis. This is proven to result when hepatocyte mechanisms for triglyceride synthesis (e.g., lipid uptake and de novo lipogenesis) overwhelm mechanisms for triglyceride disposal (e.g., degradative metabolism and lipoprotein export), leading to accumulation of fat (i.e., triglyceride) within hepatocytes. Obesity stimulates hepatocyte triglyceride accumulation by altering the intestinal microbiota to enhance both energy harvest from dietary sources and intestinal permeability. Reduced intestinal barrier function increases hepatic exposure to gut-derived products, which stimulate liver cells to generate inflammatory mediators that inhibit insulin actions. Obese adipose depots also produce excessive soluble factors (adipokines) that inhibit tissue insulin sensitivity. Insulin resistance promotes hyperglycemia, which drives the pancreas to produce more insulin to maintain glucose homeostasis. However, hyperinsulinemia also promotes lipid uptake, fat synthesis, and fat storage. The net result is hepatic triglyceride accumulation (i.e., steatosis).

Triglyceride per se is not hepatotoxic. However, its precursors (e.g., fatty acids and diacylglycerols) and metabolic by-products (e.g., reactive oxygen species) may damage hepatocytes, leading to hepatocyte lipotoxicity. Lipotoxicity also triggers the generation of other factors (e.g., inflammatory cytokines, hormonal mediators) that deregulate systems that normally maintain hepatocyte viability. The net result is increased hepatocyte death. Dying hepatocytes, in turn, release various factors that trigger wound healing responses that aim to replace (regenerate) lost hepatocytes. Such repair involves transient expansion of other cell types, such as myofibroblasts and progenitor cells, that make and degrade matrix, remodel the vasculature, and generate replacement hepatocytes, as well as the recruitment of immune cells that release factors that modulate liver injury and repair. NASH is

the morphologic manifestation of lipotoxicity and resultant wound healing responses. Because the severity and duration of lipotoxic liver injury dictate the intensity and duration of repair, the histologic features and outcomes of NASH are variable. Cirrhosis and liver cancer are potential outcomes of chronic NASH. Cirrhosis results from futile repair, i.e., progressive accumulation of wound healing cells, fibrous matrix, and abnormal vasculature (scarring), rather than efficient reconstruction/regeneration of healthy hepatic parenchyma. Primary liver cancers develop when malignantly transformed liver cells escape mechanisms that normally control regenerative growth. The mechanisms responsible for futile repair (cirrhosis) and liver carcinogenesis are not well understood. Because normal liver regeneration is a very complex process, there are multiple opportunities for deregulation and, thus, pathogenic heterogeneity. To date, this heterogeneity has confounded development of both diagnostic tests and treatments for defective/deregulated liver repair (i.e., cirrhosis and cancer). Hence, current strategies focus on circumventing misrepair by preventing and/or reducing lipotoxic liver injury.

■ DIAGNOSIS

Diagnosing NAFLD requires demonstration of increased liver fat in the absence of hazardous levels of alcohol consumption. Thresholds for potentially dangerous alcohol ingestion have been set at more than one drink per day in women and two drinks per day in men based on epidemiologic evidence that the prevalence of serum aminotransferase elevations increases when alcohol consumption habitually exceeds these levels. In those studies, one drink was defined as having 10 g of ethanol and, thus, is equivalent to one can of beer, 4 oz of wine, or 1.5 oz (one shot) of distilled spirits. Other causes of liver fat accumulation (particularly exposure to certain drugs; Table 343-2) and liver injury (e.g., viral hepatitis, autoimmune liver disease, iron or copper overload, α₁-antitrypsin deficiency) must also be excluded. Thus, establishing the diagnosis of NAFLD does not require invasive testing; it can be accomplished by history and physical examination, liver imaging (ultrasound is an acceptable first-line test; computed tomography [CT] or magnetic resonance imaging [MRI] enhances sensitivity for liver fat detection but adds expense), and blood tests to exclude other liver diseases.

It is important to emphasize that, in individuals with NAFLD, the liver may not be enlarged and serum aminotransferases and liver function tests (e.g., bilirubin, albumin, prothrombin time) may be completely normal. Because there is yet no one specific blood test for NAFLD, confidence in the diagnosis of NAFLD is increased by identification of NAFLD risk factors. The latter include increased body mass index, insulin resistance/type 2 diabetes mellitus, and other parameters indicative of the metabolic syndrome (e.g., systemic hypertension, dyslipidemia, hyperuricemia/gout, cardiovascular disease; Chap. 408) in the patient or family members. Individuals who have, or have had, pituitary or hypothalamic neoplasms and women with polycystic ovary syndrome are also at increased risk for NAFLD. Hypothyroidism and obstructive sleep apnea may also increase NAFLD, presumably by promoting obesity and/or exacerbating the metabolic syndrome.

Establishing the severity of NAFLD-related liver injury and related scarring (i.e., staging NAFLD) is more difficult than simply diagnosing NAFLD. Staging is critically important, however, because it is necessary to define prognosis and thereby determine treatment recommendations. The goal of staging is to distinguish patients with NASH from those with simple steatosis and to identify which of the NASH patients have advanced fibrosis. The 10-year probability of developing liver-related morbidity or mortality in steatosis is negligible, and hence, this subgroup of NAFLD patients tends to be managed conservatively (see below). In contrast, more intensive follow-up and therapy are justified in NASH patients, and the subgroup with advanced fibrosis merits the most intensive scrutiny and intervention because their 10-year risk of liver-related morbidity and mortality is clearly increased.

Staging approaches can be separated into noninvasive testing (i.e., blood testing, physical examination, and imaging) and invasive approaches (i.e., liver biopsy). Blood test evidence of hepatic dysfunction (e.g., hyperbilirubinemia, hypoalbuminemia, prothrombin time prolongation) or portal hypertension (e.g., thrombocytopenia) and

stigmata of portal hypertension on physical examination (e.g., spider angioma, palmar erythema, splenomegaly, ascites, clubbing, encephalopathy) suggest a diagnosis of advanced NAFLD. Liver biopsy has been the gold standard for establishing the severity of liver injury and fibrosis because it is both more sensitive and more specific than these other tests for establishing NAFLD severity. Further, although invasive, liver biopsy is seldom complicated by serious adverse sequelae such as significant bleeding, pain, or inadvertent puncture of other organs and thus is relatively safe. However, biopsy suffers from potential sampling error unless tissue cores of 2 cm or longer are acquired. Also, examination of tissue at a single point in time is not reliable for determining whether the pathologic processes are progressing or regressing. The risk of serial liver biopsies within short time intervals is generally deemed as unacceptable outside of research studies. These limitations of liver biopsy have stimulated efforts to develop noninvasive approaches to stage NAFLD.

As is true for many other types of chronic liver disease, in NAFLD, the levels of serum aminotransferases (aspartate aminotransferase [AST] and ALT) do not reliably reflect the severity of liver cell injury, extent of liver-cell death, or related liver inflammation and fibrosis. Thus, they are imperfect for determining which individuals with NAFLD have NASH. This has prompted efforts to identify superior markers of NASH and, particularly, liver fibrosis, because fibrosis stage predicts eventual liver outcomes and mortality in NASH. Algorithms that combine various laboratory tests (e.g., Enhanced Liver Fibrosis [ELF] score, BARD score, AST to Platelet Ratio Index [APRI] score, NAFLD fibrosis score, and Fibrosis-4 [FIB-4] score) are somewhat helpful in separating NASH patients with advanced versus mild liver fibrosis. The NAFLD fibrosis score (NFS) and FIB-4 score, two of the most commonly employed noninvasive tests to assess severity of hepatic fibrosis, can be calculated from a few readily available clinical variables (age, body mass index, glucose, platelet count, albumin, AST, ALT) using published formulas that are readily accessed via an online calculator. Both scores are helpful for gauging the severity of NASH and liver fibrosis. Combining these tests with new imaging approaches that permit noninvasive quantification of liver fat (e.g., MRI using proton density fat fraction [MRI-PDFF]) and liver stiffness, a surrogate marker of liver fibrosis (e.g., magnetic resonance elastography [MRE], and transient elastography [FibroScan]), improves their predictive power (Chap. 337). Transient elastography in particular has become widely available and is relatively inexpensive. It is most useful for excluding advanced liver fibrosis as cirrhosis is extremely unlikely when the liver stiffness score is low. However, higher stiffness scores must be interpreted with caution since several factors (obesity, nonfasting state, hepatic inflammation, iron overload, and/or hepatic congestion) decrease the specificity of the test. Increasingly, these new serologic and imaging tools are being used serially or in combination to monitor fibrosis progression and regression in NAFLD patients. As a result, liver biopsy staging is becoming restricted to patients who cannot be stratified reliably using these noninvasive assessments. Indeterminant or discordant results of noninvasive testing should prompt referral to a liver specialist and consideration of liver biopsy.

■ CLINICAL FEATURES OF NAFLD

Most subjects with NAFLD are asymptomatic. The diagnosis is often made when abnormal liver aminotransferases or features of fatty liver are noted during an evaluation performed for other reasons. NAFLD may also be diagnosed during the workup of vague right upper quadrant abdominal pain, hepatomegaly, or an abnormal-appearing liver at time of abdominal surgery. Obesity is present in 50–90% of subjects. Most patients with NAFLD also have other features of the metabolic syndrome (Chap. 408). Some have subtle stigmata of chronic liver disease, such as spider angioma, palmar erythema, or splenomegaly. In a small minority of patients with advanced NAFLD, complications of end-stage liver disease (e.g., jaundice, features of portal hypertension such as ascites or variceal hemorrhage) may be the initial findings.

The association of NAFLD with obesity, diabetes, hypertriglyceridemia, hypertension, and cardiovascular disease is well known. Other associations include chronic fatigue, mood alterations, obstructive

sleep apnea, thyroid dysfunction, polycystic ovary syndrome, and chronic pain syndrome. NAFLD is an independent risk factor for metabolic syndrome (Chap. 408). Longitudinal studies suggest that patients with NASH are at two- to threefold increased risk for the development of metabolic syndrome. Similarly, studies have shown that patients with NASH have a higher risk for the development of hypertension and diabetes mellitus. The presence of NAFLD is also independently associated with endothelial dysfunction, increased carotid intimal thickness, and the number of plaques in carotid and coronary arteries. Such data indicate that NAFLD has many deleterious effects on health in general.

■ TREATMENT OF NAFLD

Treatment of NAFLD can be divided into three components: (1) specific therapy of NAFLD-related liver disease; (2) treatment of NAFLD-associated comorbidities; and (3) treatment of the complications of advanced NAFLD. The subsequent discussion focuses on specific therapies for NAFLD, with some mention of their impact on major NAFLD comorbidities (insulin resistance/diabetes, obesity, and dyslipidemia). Treatment of the complications of advanced NAFLD involves management of the complications of cirrhosis and portal hypertension, including primary liver cancers. Approaches to accomplish these objectives are similar to those used in other chronic liver diseases and are covered elsewhere in the textbook (Chaps. 344 and 82).

At present, there are no U.S. Food and Drug Administration (FDA)-approved therapies for the treatment of NAFLD. Thus, the current approach to NAFLD management focuses on treatment to improve the risk factors for NASH (i.e., obesity, insulin resistance, metabolic syndrome, dyslipidemia). Based on our understanding of the natural history of NAFLD, only patients with NASH or hepatic fibrosis are considered currently for targeted pharmacologic therapies. This approach may change as our understanding of disease pathophysiology improves and potential targets of therapy evolve.

Diet and Exercise Lifestyle changes and dietary modifications that result in weight loss and/or improve insulin sensitivity are the primary treatments for NAFLD. Many studies indicate that loss of 3–5% of body weight improves steatosis and that greater weight loss (i.e., $\geq 10\%$) improves steatohepatitis and hepatic fibrosis. The benefits of modifying dietary macronutrient contents (e.g., low-carbohydrate vs low-fat diets, saturated vs unsaturated fat diets) generally parallel changes in calorie consumption, suggesting that diet modifications are mainly beneficial because they reduce energy intake and improve obesity. However, a Mediterranean-type diet has been reported to improve NASH and liver fibrosis independently of weight loss. Excluding foods and beverages high in added fructose and increasing coffee consumption are also recommended because high-fructose diets have been shown to exacerbate hepatic steatosis, steatohepatitis, and fibrosis, and consuming two or more cups of coffee per day is associated with reduced risk of liver fibrosis. Changes in diet composition particularly merit consideration in lean individuals with NAFLD, although available data are insufficient to determine if this improves their liver histology. Modifying lifestyle to increase physical activity (i.e., energy expenditure) complements dietary calorie restriction and, thus, expedites weight loss. Exercise also improves muscle insulin sensitivity, which improves the metabolic syndrome independent of weight loss. Both aerobic exercise and resistance training effectively reduce liver fat. At least 30 min of moderate-intensity aerobic exercise or resistance training five times per week is recommended. The choice of training should be tailored to patients' preferences and functional capacity to enable long-term maintenance. Any activity is better than remaining sedentary. Unfortunately, most NAFLD patients cannot sustain long-term compliance with diet and lifestyle modifications and, thus, fail to maintain a healthier weight. Although pharmacologic therapies to facilitate weight loss, such as orlistat, topiramate, phentermine, and GLP-1 receptor agonists, are available, their role in the treatment of NAFLD remains experimental.

Pharmacologic Therapies Several drug therapies have been tried in both research and clinical settings. There are currently no FDA-approved drugs for the treatment of NAFLD. Hence, at present,

NAFLD patients without NASH or fibrosis should receive only counseling for healthy diet and physical activity. Consideration of additional specific pharmacotherapy for liver disease is restricted to NAFLD patients with more serious liver damage (i.e., NASH or liver fibrosis). A number of large clinical trials designed to identify effective and safe treatments for these conditions are in progress. Because NAFLD is strongly associated with the metabolic syndrome and type 2 diabetes (*Chaps. 403 and 404*), the efficacy of various insulin-sensitizing agents has been examined. *Metformin*, an agent that mainly improves hepatic insulin sensitivity, has been evaluated in several small, open-label studies in adults and a recent larger, prospectively randomized trial in children (dubbed the TONIC study). Although several of the adult NASH studies suggested improvements in aminotransferases and, less consistently, liver histology, metformin did not improve liver histology in the TONIC study of children with NASH. Thus, it is not currently recommended as a treatment for NASH. *Thiazolidinediones* (*pioglitazone and rosiglitazone*), drugs known to improve systemic insulin resistance, have been studied in adults with NASH. Both agents reduced aminotransferases and improved some of the histologic features of NASH in small, uncontrolled studies. A large, randomized, placebo-controlled clinical trial sponsored by the National Institutes of Health, the PIVENS Study (*Pioglitazone vs Vitamin E vs Placebo for the Treatment of 247 Nondiabetic Adults with NASH*), demonstrated that resolution of histologic NASH occurred more often in subjects treated with *pioglitazone* (30 mg/d) than with placebo for 18 months (47 vs 21%, $p = .001$). However, many subjects in the pioglitazone group gained weight, and liver fibrosis did not improve. Five-year follow-up of subjects who were treated with *rosiglitazone* for up to 2 years demonstrated that extending treatment and follow-up duration did not further improve NASH or liver fibrosis, and rosiglitazone has been associated with increased long-term risk for cardiovascular mortality. Pioglitazone may be safer than rosiglitazone, however, because in a recent large meta-analysis it was associated with reduced overall morality, myocardial infarction, and stroke. Caution is still warranted, however, because long-term use of thiazolidinediones has been associated with weight gain, increased risk for bladder cancer, and bone fractures in women.

Incretin mimetics, drugs that act on the pancreas to optimize insulin and glucagon release, have improved liver enzyme elevations. A small pilot trial of daily injections of *liraglutide* and phase 2 studies of *semaglutide* demonstrated remission of NASH without worsening liver fibrosis. Agents that improve hyperglycemia by blocking renal reabsorption of glucose, *sodium glucose cotransporter (SGLT2 inhibitors*, have been observed to improve serum liver enzymes in diabetic patients with and are also under formal evaluation as treatments for NASH. Both *incretin mimetics* and *SGLT2 inhibitors* can be used in NASH patients with type 2 diabetes or obesity (conditions for which the drugs have an FDA-registered indication for use); however, they are not currently approved specifically for the treatment of NASH.

Antioxidants have also been evaluated for the treatment of NAFLD because oxidant stress is thought to contribute to the pathogenesis of NASH. *Vitamin E*, an inexpensive yet potent antioxidant, has been examined in several small pediatric and adult studies with varying results. In all of those studies, vitamin E was well tolerated, and most studies showed modest improvements in aminotransferase levels, radiographic features of hepatic steatosis, and/or histologic features of NASH. Vitamin E (800 IU/d) was compared to placebo in the PIVENS and TONIC studies. In PIVENS, vitamin E was the only agent that achieved the predetermined primary endpoint (i.e., improvement in steatohepatitis without worsening of fibrosis). This endpoint was met in 43% of patients in the vitamin E group ($p = .001$ vs placebo), 34% in the pioglitazone group ($p = .04$ vs placebo), and 19% in the placebo group. Vitamin E also improved NASH histology in pediatric patients with NASH (TONIC trial). However, recent population-based studies suggest that chronic vitamin E therapy may increase the risk for cardiovascular mortality, hemorrhagic stroke, and prostate cancer. Thus, vitamin E should only be considered as a first-line pharmacotherapy for nondiabetic, noncirrhotic NASH patients who are at low risk for cardiovascular disease or prostate cancer. Further studies are needed before firm recommendations can be made regarding the

risk-to-benefit ratio and long-term therapeutic efficacy of vitamin E in NASH. *Ursodeoxycholic acid* (a bile acid that improves certain cholestatic liver diseases) and *betaine* (a metabolite of choline that raises *S*-adenosylmethionine [SAM] levels and decreases cellular oxidative damage) offer no histologic benefit over placebo in patients with NASH. Experimental evidence to support the use of *omega-3 fatty acids* in NAFLD exists; however, a recent large, multicenter, placebo-controlled study failed to demonstrate a histologic benefit.

Many other pharmacotherapies that target dysregulated energy homeostasis, lipotoxicity, cell death, and liver inflammation, processes that are critically involved in the pathogenesis and/or progression of NASH and liver fibrosis, are currently in clinical trials (e.g., *probiotics, farnesoid X receptor agonists, fibroblast growth factor agonists, anti-apoptotic agents, anticytokine agents, dipeptidyl IV antagonists, PPAR modulators, thyroid hormone receptor β-selective agonists, stearly-CoA desaturase-1 inhibitors, DGAT inhibitors, acyl-CoA carboxylase inhibitors, and direct modulators of liver fibrosis*). Sufficient data do not yet exist to justify their use as NASH treatments in clinical practice. Given that liver disease outcomes in NASH patients are highly heterogeneous, optimal treatment of NASH may need to be individualized by tailoring therapy based on clinical or histologic phenotypes of NASH and/or genetic susceptibility for disease progression.

Statins are an important class of agents to treat dyslipidemia and decrease cardiovascular risk. There is no evidence to suggest that statins cause liver failure in patients with any chronic liver disease, including NAFLD. The incidence of liver enzyme elevations in NAFLD patients taking statins is also no different than that of healthy controls or patients with other chronic liver diseases. Moreover, several studies have suggested that statins may improve aminotransferases and histology in patients with NASH. Yet there is continued reluctance to use statins in patients with NAFLD. The lack of evidence that statins harm the liver in NAFLD patients, combined with the increase risk for cardiovascular morbidity and mortality in NAFLD patients, justifies the use of statins to treat dyslipidemia in patients with NAFLD/NASH.

Bariatric Surgery Although interest in bariatric surgery as a treatment for NAFLD exists, a recently published Cochrane review concluded that lack of randomized clinical trials or adequate clinical studies prevents definitive assessment of benefits and harms of bariatric surgery as a treatment for NASH. Most studies of bariatric surgery have shown that bariatric surgery is generally safe in individuals with well-compensated chronic liver disease and improves hepatic steatosis and necroinflammation (i.e., features of NAFLD/NASH); however, effects on hepatic fibrosis have been variable. Concern lingers because some of the largest prospective studies suggest that hepatic fibrosis might progress after bariatric surgery. Thus, the Cochrane review deemed it premature to recommend bariatric surgery as a primary treatment for NASH. This opinion was challenged by a recently study that demonstrated that fibrosis stage had improved by 5 years after surgery in about half the patients in one large bariatric surgery cohort. However, most of those individuals had relatively mild fibrosis initially, and thus, it is unclear if similar outcomes would occur in individuals with more advanced liver disease. Indeed, there is general agreement that patients with NAFLD-related cirrhosis and, particularly, those with portal hypertension should be excluded as candidates for bariatric surgery. However, given growing evidence for the benefits of bariatric surgery on metabolic syndrome complications in individuals with refractory obesity, it is not contraindicated in otherwise eligible patients with NAFLD or NASH.

Liver Transplantation Patients with NAFLD in whom end-stage liver disease develops should be evaluated for liver transplantation (*Chap. 345*). The outcomes of liver transplantation in well-selected patients with NAFLD are generally good, but comorbid medical conditions associated with NAFLD, such as diabetes mellitus, obesity, and cardiovascular disease, often limit transplant candidacy. NAFLD may recur after liver transplantation. The risk factors for recurrent or de novo NAFLD after liver transplantation are multifactorial and include hypertriglyceridemia, obesity, diabetes mellitus, and immunosuppressive therapies, particularly glucocorticoids.

Obesity is an accelerating global disease. The worldwide prevalence of obesity has more than doubled since 1980, and there are now >1 billion overweight adults, of whom at least 300 million are obese. In the wake of the obesity epidemic follow numerous comorbidities, including NAFLD. NAFLD is the most common liver disease identified in Western countries and the fastest rising form of chronic liver disease worldwide. The economic burden directly attributable to NAFLD is already enormous (estimated direct medical costs of ~\$103 billion/year in the United States and €35 billion/year in the Europe-4 countries: Germany, France, Italy, and United Kingdom) and predicted to increase tenfold by the year 2025. Present understanding of NAFLD's natural history is based mainly on studies in whites who became overweight/obese and developed the metabolic syndrome in adulthood. The impact of the global childhood obesity epidemic on NAFLD pathogenesis/progression is unknown. Emerging evidence demonstrates that advanced NAFLD, including cirrhosis and primary liver cancer, can occur in children, prompting concerns that childhood-onset NAFLD might follow a more aggressive course than typical adult-acquired NAFLD. Some of the most populated parts of the world are in the midst of industrial revolutions, and certain environmental pollutants seem to exacerbate NAFLD. Some studies also suggest that the risk for NASH and NAFLD-related cirrhosis may be higher in certain ethnic groups such as Asians, Hispanics, and Native Americans, and lower in others such as African Americans, compared with whites. Although all of these variables confound efforts to predict the net impact of this obesity-related liver disease on global health, it seems likely that NAFLD will remain a major cause of chronic liver disease worldwide for the foreseeable future.

■ FURTHER READING

- C N et al: The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 67:328, 2018.
- D AM, D CSC: Cause, pathogenesis, and treatment of nonalcoholic steatohepatitis. *N Engl J Med* 377:2063, 2017.
- E A S L (EASL) et al: EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 64:1388, 2016.
- V MB et al: NASPGHAN clinical practice guideline for the diagnosis and treatment of nonalcoholic fatty liver disease in children: Recommendations from the Expert Committee on NAFLD (ECON) and the North American Society of Pediatric Gastroenterology, Hepatology and Nutrition (NASPGHAN). *J Pediatr Gastroenterol Nutr* 64:319, 2017.

TABLE 344-1 Causes of Cirrhosis

Alcohol	Cardiac cirrhosis
Chronic viral hepatitis	Inherited metabolic liver disease
Hepatitis B	Hemochromatosis
Hepatitis C	Wilson's disease
Autoimmune hepatitis	α ₁ Antitrypsin deficiency
Nonalcoholic steatohepatitis	Cystic fibrosis
Biliary cirrhosis	Cryptogenic cirrhosis
Primary biliary cholangitis	
Primary sclerosing cholangitis	
Autoimmune cholangiolopathy	

distortion with the formation of regenerative nodules. This results in a decrease in hepatocellular mass, and thus function, and an alteration of blood flow. The induction of fibrosis occurs with activation of hepatic stellate cells, resulting in the formation of increased amounts of collagen and other components of the extracellular matrix.

Clinical features of cirrhosis are the result of pathologic changes and mirror the severity of the liver disease. Most hepatic pathologists provide an assessment of grading and staging when evaluating liver biopsy samples. These grading and staging schemes vary between disease states and have been developed for most conditions, including chronic viral hepatitis, nonalcoholic fatty liver disease, and primary biliary cholangitis. Advanced fibrosis usually includes bridging fibrosis with nodularity designated as stage 3 and cirrhosis designated as stage 4. Patients who have cirrhosis have varying degrees of liver function, and clinicians need to differentiate between those who have stable, compensated cirrhosis and those who have decompensated cirrhosis. Patients who have developed ascites, hepatic encephalopathy, or variceal bleeding are classified as decompensated. They should be considered for liver transplantation, particularly if the decompensations are poorly controlled. Many of the complications of cirrhosis will require specific therapy. *Portal hypertension* is a significant complicating feature of decompensated cirrhosis and is responsible for the development of ascites and bleeding from esophagogastric varices, two complications that signify decompensated cirrhosis. Loss of hepatocellular function results in jaundice, coagulation disorders, and hypoalbuminemia and contributes to the causes of portosystemic encephalopathy. The complications of cirrhosis are basically the same regardless of the etiology. Nonetheless, it is useful to classify patients by the cause of their liver disease (Table 344-1); patients can be divided into broad groups, including those with alcohol-associated cirrhosis, cirrhosis due to chronic viral hepatitis, biliary cirrhosis, nonalcoholic fatty liver disease, and other, less common causes, such as cardiac cirrhosis, cryptogenic cirrhosis, and other miscellaneous causes.

ALCOHOL ASSOCIATED CIRRHOSIS

Excessive chronic alcohol use can cause several different types of chronic liver disease, including alcohol-associated fatty liver, alcoholic hepatitis, and alcohol-associated cirrhosis. Furthermore, use of excessive alcohol can contribute to liver damage in patients with other liver diseases, such as hepatitis C, hemochromatosis, and fatty liver disease related to obesity. Chronic alcohol use can produce fibrosis in the absence of accompanying inflammation and/or necrosis. Fibrosis can be centrilobular, pericellular, or periportal. When fibrosis reaches a certain degree, there is disruption of the normal liver architecture and replacement of liver cells by regenerative nodules. In alcohol-associated cirrhosis, the nodules are usually <3 mm in diameter; this form of cirrhosis is referred to as *micronodular*. With cessation of alcohol use, larger nodules may form, resulting in a mixed micronodular and macronodular cirrhosis.

Pathogenesis Alcohol is the most commonly used drug in the United States, and >70% of adults drink alcohol each year. Twenty percent have had a binge within the past month, and >7% of adults regularly consume more than four or five drinks five or more times a month. Unfortunately, >14 million adults in the United States meet

344

Cirrhosis and Its Complications

Alex S. Befeler, Bruce R. Bacon

Cirrhosis is a condition that is defined histopathologically and has a variety of clinical manifestations and complications, some of which can be life-threatening. In the past, it has been thought that cirrhosis was never reversible; however, it has become apparent that when the underlying insult that has caused the cirrhosis has been removed, there can be reversal of fibrosis. This is most apparent with the successful treatment of chronic hepatitis C; however, reversal of fibrosis is also seen in patients with hemochromatosis who have been successfully treated and in patients with alcohol associate liver disease who have discontinued alcohol use.

Regardless of the cause of cirrhosis, the pathologic features consist of the development of fibrosis to the point that there is architectural

the diagnostic criteria for alcohol use disorder. In the United States, chronic liver disease is the tenth most common cause of death in adults, and alcohol-associated cirrhosis accounts for ~48% of deaths due to cirrhosis.

Ethanol is mainly absorbed by the small intestine and, to a lesser degree, through the stomach. Gastric alcohol dehydrogenase (ADH) initiates alcohol metabolism. Three enzyme systems account for metabolism of alcohol in the liver. These include cytosolic ADH, the microsomal ethanol oxidizing system (MEOS) utilizing the inducible cytochrome P450 CYP2E1, and peroxisomal catalase. Normally the majority of ethanol oxidation occurs via ADH to form acetaldehyde, which is a highly reactive molecule that may have multiple effects. The MEOS pathway in chronic alcohol use causes induction of CYP2E1, which leads to generation of reactive oxygen species and produces more acetaldehyde. Ultimately, acetaldehyde is metabolized to acetate by aldehyde dehydrogenase (ALDH). Intake of ethanol increases intracellular accumulation of triglycerides by increasing fatty acid uptake and by reducing fatty acid oxidation and lipoprotein secretion. Protein synthesis, glycosylation, and secretion are impaired. Oxidative damage to hepatocyte membranes occurs due to the formation of reactive oxygen species; acetaldehyde is a highly reactive molecule that combines with proteins and nucleic acids to form acetaldehyde adducts. These adducts may interfere with specific enzyme activities, including microtubular formation and hepatic protein trafficking. With acetaldehyde-mediated hepatocyte damage, certain reactive oxygen species can result in Kupffer cell activation. As a result, profibrogenic cytokines are produced that initiate and perpetuate stellate cell activation, with the resultant production of excess collagen and extracellular matrix. Connective tissue appears in both periportal and pericentral zones and eventually connects portal triads with central veins forming regenerative nodules. Hepatocyte loss occurs, and with increased collagen production and deposition, together with continuing hepatocyte destruction, the liver contracts and shrinks in size. This process generally takes from years to decades to occur and requires repeated insults.

Clinical Features The diagnosis of alcohol associate liver disease requires an accurate history regarding both amount and duration of alcohol consumption. Patients with alcohol associate liver disease can present with nonspecific symptoms such as vague right upper quadrant abdominal pain, fever, nausea and vomiting, diarrhea, anorexia, and malaise. Alternatively, they may present with more specific complications of chronic liver disease, including ascites, edema, upper gastrointestinal (GI) hemorrhage, jaundice, or encephalopathy. Many cases present incidentally at the time of autopsy or elective surgery. The abrupt onset of any of these complications may be the first event prompting the patient to seek medical attention. Other patients may be identified in the course of an evaluation of routine laboratory studies that are found to be abnormal. On physical examination, the liver and spleen may be enlarged, with the liver edge being firm and nodular. Other frequent findings include scleral icterus, palmar erythema (Fig. 344-1), spider angiomas (Fig. 344-2),



FIGURE 344-1 Palmar erythema. This figure shows palmar erythema in a patient with alcohol-associated cirrhosis. The erythema is peripheral over the palm with central pallor.

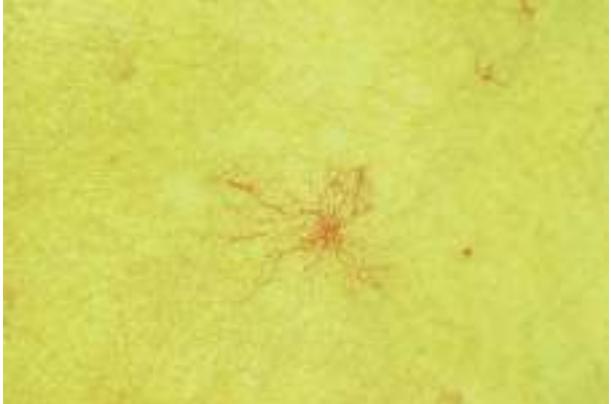


FIGURE 344-2 Spider angioma. This figure shows a spider angioma in a patient with hepatitis C cirrhosis. With release of central compression, the arteriole fills from the center and spreads out peripherally.

parotid gland enlargement, digital clubbing, muscle wasting, edema, and ascites. Men may have decreased body hair and gynecomastia as well as testicular atrophy, which may be a consequence of hormonal abnormalities or a direct toxic effect of alcohol on the testes. In women with advanced alcohol-associated cirrhosis, menstrual irregularities usually occur including amenorrhea. These changes are often reversible following cessation of alcohol ingestion.

Laboratory tests may be completely normal in patients with early compensated alcohol-associated cirrhosis. Alternatively, in advanced liver disease, many abnormalities usually are present. Patients may be anemic from chronic GI blood loss, nutritional deficiencies, or hypersplenism or as a direct suppressive effect of alcohol on the bone marrow. A unique form of hemolytic anemia (with spur cells and acanthocytes) called *Zieve's syndrome* can occur in patients with severe alcoholic hepatitis. Platelet counts are often reduced early in the disease, reflective of portal hypertension with hypersplenism. Serum total bilirubin can be normal or elevated with advanced disease. Prothrombin times are often prolonged and usually do not respond to administration of parenteral vitamin K. Serum sodium levels are usually normal unless patients have ascites and then can be depressed, largely due to ingestion of excess free water. Serum alanine and aspartate aminotransferases (ALT, AST) are typically elevated, particularly in patients who continue to drink, with AST levels being higher than ALT levels, usually by a 2:1 ratio.

Diagnosis Patients who have any of the above-mentioned clinical features, physical examination findings, or laboratory studies should be considered to have alcohol associate liver disease. The diagnosis, however, requires accurate knowledge that the patient is continuing to use or has recently stopped alcohol. Furthermore, other forms of chronic liver disease (e.g., chronic viral hepatitis or metabolic or autoimmune liver diseases) must be considered or ruled out, or if present, an estimate of relative causality along with the alcohol use should be determined. Liver biopsy can be helpful to confirm a diagnosis but generally is not performed unless there is a suspicion of an alternative diagnosis.

In patients who have had complications of cirrhosis and who continue to drink, there is a <50% 5-year survival. In contrast, in patients who are able to remain abstinent, the prognosis is significantly improved, particularly when they have resolution of liver complications; however, some individuals who remain abstinent do not improve and liver transplantation is a viable option.

TREATMENT

Alcohol-Associated Cirrhosis and Alcoholic Hepatitis

Abstinence is the cornerstone of therapy for patients with alcohol associate liver disease. In addition, patients require good nutrition and long-term medical supervision to manage underlying complications that may develop. Complications such as the development of

ascites and edema, variceal hemorrhage, or portosystemic encephalopathy all require specific management and treatment. Liver transplantation can be an effective long-term treatment in those who have been deemed a low enough risk for alcohol relapse and do not respond to other treatments.

Glucocorticoids are occasionally used in patients with severe alcoholic hepatitis in the absence of infection. Short-term survival has been shown to be improved in certain studies and meta-analysis, although 6-month survival is more dependent on abstinence. Treatment is restricted to patients with a discriminant function (DF) value of >32 . The DF is calculated as the serum total bilirubin plus the difference in the patient's prothrombin time compared to upper limit of control (in seconds) multiplied by 4.6. Failure to improve total bilirubin after 7 days predicts treatment failure, and glucocorticoids can be stopped; otherwise, they are continued for 28 days.

There is modest evidence that intravenous *N*-acetylcysteine plus glucocorticoids may have survival benefit in alcoholic hepatitis if the DF is >32 . Other therapies including oral pentoxifylline, parenterally administered inhibitors of tumor necrosis factor (TNF) α such as infliximab or etanercept, anabolic steroids, propylthiouracil, antioxidants, colchicine, and penicillamine have not shown clear-cut benefits and are not recommended. A variety of nutritional therapies have been tried, both parenteral and enteral feedings; however, there is no clear evidence of improved survival. There is evidence that persons who consume >21.5 kcal/kg body weight per day have better survival, so achieving better caloric intake is recommended. Finally, in highly selected patients with good social support structure who fail other treatments for alcoholic hepatitis, early liver transplant can be an effective treatment.

The cornerstone to treatment is cessation of alcohol use. Recent experience with medications that reduce craving for alcohol, such as acamprosate calcium and baclofen, have been favorable. Patients may take other necessary medications even in the presence of cirrhosis. Acetaminophen use is often discouraged in patients with liver disease; however, if no more than 2 g of acetaminophen per day are consumed, there generally are no problems unless there is active alcohol use.

CIRRHOSIS DUE TO CHRONIC VIRAL HEPATITIS B OR C

Of patients exposed to the hepatitis C virus (HCV), $\sim 80\%$ develop chronic hepatitis C, and of those, ~ 20 – 30% will develop cirrhosis over 20–30 years. Many of these patients have had concomitant alcohol use, and the true incidence of cirrhosis due to hepatitis C alone is unknown. It is expected that an even higher percentage will go on to develop cirrhosis over longer periods of time. In the United States, ~ 5 – 6 million people have been exposed to HCV, and ~ 4 – 5 million are chronically viremic. Worldwide, ~ 170 million individuals have hepatitis C, with some areas of the world (e.g., Egypt) having up to 15% of the population infected. HCV is a noncytopathic virus, and liver damage is probably immune-mediated. Progression of liver disease due to chronic hepatitis C is characterized by portal-based fibrosis with bridging fibrosis and nodularity developing, ultimately culminating in the development of cirrhosis. In cirrhosis due to chronic hepatitis C, the liver is small and shrunken with characteristic features of a mixed micro- and macronodular cirrhosis seen on liver biopsy. In addition to the increased fibrosis that is seen in cirrhosis due to hepatitis C, an inflammatory infiltrate is found in portal areas with interface hepatitis and occasionally some lobular hepatocellular injury and inflammation. In patients with HCV genotype 3, steatosis is often present.

Similar findings are seen in patients with cirrhosis due to chronic hepatitis B. Of adult patients exposed to hepatitis B, $\sim 5\%$ develop chronic hepatitis B, and $\sim 20\%$ of those patients will go on to develop cirrhosis. Special stains for hepatitis B core (HBc) and hepatitis B surface (HBs) antigen will be positive, and ground-glass hepatocytes signifying HBs antigen (HBsAg) may be present. In the United States, there are ~ 2 million carriers of hepatitis B, whereas in other parts of the world where hepatitis B virus (HBV) is endemic (i.e., Asia,

Southeast Asia, sub-Saharan Africa), up to 15% of the population may be infected, having acquired the infection vertically at the time of birth. Thus, >300 – 400 million individuals are thought to have hepatitis B worldwide. Approximately 25% of these individuals may ultimately develop cirrhosis.

Clinical Features and Diagnosis Patients with cirrhosis due to either chronic hepatitis C or B can present with the usual symptoms and signs of chronic liver disease. Fatigue, malaise, vague right upper quadrant pain, and laboratory abnormalities are frequent presenting features. Diagnosis requires a thorough laboratory evaluation, including quantitative HCV RNA testing and analysis for HCV genotype, or hepatitis B serologies to include HBsAg, anti-HBs, HBeAg (hepatitis B e antigen), anti-HBe, and quantitative HBV DNA levels.

TREATMENT

Cirrhosis due to Chronic Viral Hepatitis B or C

Management of complications of cirrhosis revolves around specific therapy for treatment of whatever complications occur (e.g., esophageal variceal hemorrhage, development of ascites and edema, or encephalopathy). In patients with chronic hepatitis B, numerous studies have shown beneficial effects of antiviral therapy, which is effective at viral suppression, as evidenced by reducing aminotransferase levels and HBV DNA levels and improving histology by reducing inflammation and fibrosis. Several clinical trials and case series have demonstrated that patients with decompensated liver disease can become compensated with the use of antiviral therapy directed against hepatitis B. Currently available therapy includes lamivudine, adefovir, telbivudine, entecavir, and tenofovir, with the latter two being preferred because of reduced risk of viral resistance. Interferon α can also be used for treating hepatitis B, but it should not be used in cirrhotics (see Chap. 341).

Treatment of patients with cirrhosis due to hepatitis C used to be more difficult because the side effects of pegylated interferon and ribavirin therapy were difficult to manage. Over the past several years, interferon-based regimens have been replaced by direct-acting antiviral protocols that are highly successful ($>95\%$ cure rate), well tolerated, and usually of short duration (8–12 weeks), but costly. These medications have truly revolutionized the treatment of hepatitis C (see Chap. 341).

CIRRHOSIS FROM AUTOIMMUNE HEPATITIS AND NONALCOHOLIC FATTY LIVER DISEASE

Other causes of posthepatitic cirrhosis include autoimmune hepatitis (AIH) and cirrhosis due to nonalcoholic steatohepatitis. Many patients with AIH present with cirrhosis that is already established. Typically, these patients will not benefit from immunosuppressive therapy with glucocorticoids or azathioprine because the AIH is “burned out.” In this situation, liver biopsy does not show a significant inflammatory infiltrate. Diagnosis in this setting requires positive autoimmune markers such as antinuclear antibody (ANA) or anti-smooth-muscle antibody (ASMA). When patients with AIH present with cirrhosis and active inflammation accompanied by elevated liver enzymes, there can be considerable benefit from the use of immunosuppressive therapy.

Patients with nonalcoholic steatohepatitis are increasingly being found to have progressed to cirrhosis. With the epidemic of obesity that continues in Western countries, more and more patients are identified with nonalcoholic fatty liver disease (Chap. 343). Of these, a significant subset has nonalcoholic steatohepatitis and can progress to increased fibrosis and cirrhosis. Over the past several years, it has been increasingly recognized that many patients who were thought to have cryptogenic cirrhosis in fact have nonalcoholic steatohepatitis. As their cirrhosis progresses, they become catabolic and then lose the telltale signs of steatosis seen on biopsy. Management of complications of cirrhosis due to either AIH or nonalcoholic steatohepatitis is similar to that for other forms of cirrhosis.

BILIARY CIRRHOSIS

Biliary cirrhosis has pathologic features that are different from either alcohol-associated cirrhosis or posthepatitic cirrhosis, yet the manifestations of end-stage liver disease are the same. Cholestatic liver disease may result from necroinflammatory lesions, congenital or metabolic processes, or external bile duct compression. Thus, two broad categories reflect the anatomic sites of abnormal bile retention: *intrahepatic* and *extrahepatic*. The distinction is important for obvious therapeutic reasons. Extrahepatic obstruction may benefit from surgical or endoscopic biliary tract decompression, whereas intrahepatic cholestatic processes will not improve with such interventions and require a different approach.

The major causes of chronic cholestatic syndromes are primary biliary cholangitis (PBC), autoimmune cholangitis (AIC), primary sclerosing cholangitis (PSC), and idiopathic adulthood ductopenia. These syndromes are usually clinically distinguished from each other by antibody testing, cholangiographic findings, and clinical presentation. However, they all share the histopathologic features of chronic cholestasis, such as cholate stasis; copper deposition; xanthomatous transformation of hepatocytes; and irregular, so-called biliary fibrosis. In addition, there may be chronic portal inflammation, interface activity, and chronic lobular inflammation. Ductopenia is a result of this progressive disease as patients develop cirrhosis.

PRIMARY BILIARY CHOLANGITIS

PBC is seen in about 100–200 individuals per million, with a strong female preponderance and a median age of ~50 years at the time of diagnosis. The cause of PBC is unknown; it is characterized by portal inflammation and necrosis of cholangiocytes in small- and medium-sized bile ducts. Cholestatic features prevail, and biliary cirrhosis is characterized by an elevated bilirubin level and progressive liver failure. Liver transplantation is the treatment of choice for patients with decompensated cirrhosis due to PBC. Ursodeoxycholic acid (UDCA) is the first-line treatment that has some degree of efficacy by slowing the rate of progression of the disease.

Antimitochondrial antibodies (AMAs) are present in ~95% of patients with PBC. These autoantibodies recognize lipoic acid on the inner mitochondrial membrane proteins that are enzymes of the pyruvate dehydrogenase complex (PDC), the branched-chain 2-oxoacid dehydrogenase complex, and the 2-oxoglutarate dehydrogenase complex. These autoantibodies are not pathogenic, but rather are useful markers for making a diagnosis.

Pathology Histopathologic analyses of liver biopsies of patients with PBC have resulted in identifying four distinct stages of the disease as it progresses. The earliest lesion is termed *chronic nonsuppurative destructive cholangitis* and is a necrotizing inflammatory process of the portal tracts. Medium and small bile ducts are infiltrated with lymphocytes and undergo duct destruction. Mild fibrosis and sometimes bile stasis can occur. With progression, the inflammatory infiltrate becomes less prominent, but the number of bile ducts is reduced and there is proliferation of smaller bile ductules. Increased fibrosis ensues with the expansion of periportal fibrosis to bridging fibrosis. Finally, cirrhosis, which may be micronodular or macronodular, develops.

Clinical Features Currently, most patients with PBC are middle-aged women diagnosed well before the end-stage manifestations of the disease are present, and as such, most patients are asymptomatic. When symptoms are present, they most prominently include a significant degree of fatigue out of proportion to either the severity of the liver disease or the age of the patient. Pruritus is seen in ~50% of patients at the time of diagnosis, and it can be debilitating. It might be intermittent and usually is most bothersome in the evening. In some patients, pruritus can develop toward the end of pregnancy and can be mistaken for cholestasis of pregnancy. Pruritus that presents prior to the development of jaundice indicates severe disease and a poor prognosis.

Physical examination can show jaundice and other complications of chronic liver disease including hepatomegaly, splenomegaly, ascites, and edema. Other features that are unique to PBC include hyperpigmentation, xanthelasma, and xanthomata, which are related to altered

cholesterol metabolism. Hyperpigmentation is evident on the trunk and the arms and is seen in areas of exfoliation and lichenification associated with progressive scratching related to the pruritus. Bone pain resulting from osteopenia or osteoporosis is occasionally seen at diagnosis.

Laboratory Findings Laboratory findings in PBC show cholestatic liver enzyme abnormalities with an elevation in γ -glutamyl transpeptidase and alkaline phosphatase (ALP) along with mild elevations in aminotransferases (ALT and AST). Immunoglobulins, particularly IgM, are typically increased. Hyperbilirubinemia usually is seen once cirrhosis has developed. Thrombocytopenia, leukopenia, and anemia may be seen in patients with portal hypertension and hypersplenism. Liver biopsy shows characteristic features as described above and should be evident to any experienced hepatopathologist. Up to 10% of patients with characteristic PBC will have features of AIH (moderate to severe interphase hepatitis on biopsy, elevated ALT >5x the upper limit of normal, and elevated IgG levels) as well and are defined as having “overlap” syndrome. These patients are usually treated as PBC patients and may progress to cirrhosis with the same frequency as typical PBC patients. Some patients require immunosuppressive medications as well.

Diagnosis PBC should be considered in patients with chronic cholestatic liver enzyme abnormalities. AMA testing may be negative in as many as 5–10% of patients with PBC. These patients usually are positive for other PBC-specific autoantibodies including sp100 or gp210, although these tests are not universally available. Liver biopsy is most important in this setting of AMA-negative PBC. In patients who are AMA negative with cholestatic liver enzymes, PSC should be ruled out by way of cholangiography.

TREATMENT

Primary Biliary Cholangitis

Treatment of the typical manifestations of cirrhosis is no different for PBC than for other forms of cirrhosis. UDCA has been shown to improve both biochemical and histologic features of the disease, thus slowing but not reversing or curing the disease. Improvement is greatest when therapy is initiated early; the likelihood of significant improvement with UDCA is low in patients with PBC who present with manifestations of cirrhosis. UDCA is given in doses of 13–15 mg/kg per d; the medication is usually well tolerated, although some patients have worsening pruritus with initiation of therapy. A small proportion of patients may have diarrhea or headache as a side effect of the drug. About 30–40% of patients with PBC do not have a satisfactory response to UDCA; about half of these patients will have significant improvement with obeticholic acid. Patients with PBC require long-term follow-up by a physician experienced with the disease. Certain patients may need to be considered for liver transplantation should their liver disease decompensate.

The main symptoms of PBC are fatigue and pruritus, and symptom management is important. Several therapies have been tried for treatment of fatigue, but none of them has been successful; frequent naps should be encouraged. Pruritus is treated with antihistamines, narcotic receptor antagonists (naltrexone), and rifampin. Cholestyramine, a bile salt-sequestering agent, has been helpful in some patients but is somewhat tedious and difficult to take. Plasmapheresis has been used rarely in patients with severe intractable pruritus. There is an increased incidence of osteopenia and osteoporosis in patients with cholestatic liver disease, and bone density testing should be performed. Oral calcium and vitamin D are also recommended. Treatment with a bisphosphonate should be instituted when bone disease is identified.

PRIMARY SCLEROSING CHOLANGITIS

As in PBC, the cause of PSC remains unknown. PSC is a chronic cholestatic syndrome that is characterized by diffuse inflammation and fibrosis involving the entire biliary tree, resulting in chronic cholestasis.

This pathologic process ultimately results in obliteration of both the intra- and extrahepatic biliary tree, leading to biliary cirrhosis, portal hypertension, and liver failure. The cause of PSC remains unknown despite extensive investigation into various mechanisms related to bacterial and viral infections, toxins, genetic predisposition, and immunologic mechanisms, all of which have been postulated to contribute to the pathogenesis and progression of this syndrome.

Liver biopsy changes in PSC are not pathognomonic, and establishing the diagnosis of PSC must involve imaging of the biliary tree. Pathologic changes occurring in PSC show bile duct proliferation as well as ductopenia and fibrous cholangitis (pericholangitis). Periductal fibrosis is occasionally seen on biopsy specimens and can be quite helpful in making the diagnosis. As the disease progresses, biliary cirrhosis is the end-stage manifestation of PSC.

Clinical Features The usual clinical features of PSC are those found in cholestatic liver disease, with fatigue, pruritus, steatorrhea, deficiencies of fat-soluble vitamins, and the associated consequences. As in PBC, the fatigue is profound and nonspecific. Pruritus can often be debilitating and is related to the cholestasis. The severity of pruritus does not correlate with the severity of the disease. Metabolic bone disease, as seen in PBC, can occur with PSC and should be treated (see above).

Laboratory Findings Patients with PSC typically are identified during an evaluation of abnormal liver enzymes. Most patients have at least a twofold increase in ALP and may have elevated aminotransferases as well. Albumin levels may be decreased, and prothrombin times are prolonged in a substantial proportion of patients at the time of diagnosis. Some degree of correction of a prolonged prothrombin time may occur with parenteral vitamin K. A small subset of patients has aminotransferase elevations >5 times the upper limit of normal and may have features of AIH on biopsy indicating an overlap syndrome between PSC and AIH. Autoantibodies are frequently positive in patients with the overlap syndrome but are typically negative in patients who only have PSC. One autoantibody, the perinuclear anti-neutrophil cytoplasmic antibody (pANCA), is positive in ~65% of patients with PSC. Sixty to eighty percent of patients with PSC have inflammatory bowel disease, predominately ulcerative colitis (UC); thus, a colonoscopy is recommended at diagnosis.

Diagnosis The definitive diagnosis of PSC requires cholangiographic imaging. Over the past several years, magnetic resonance imaging (MRI) with magnetic resonance cholangiopancreatography (MRCP) has been used as the imaging technique of choice for initial evaluation. Endoscopic retrograde cholangiopancreatography (ERCP) should be performed if the MRCP provided suboptimal images or if there is clinical (newly elevated total bilirubin or worsening pruritus) or MRCP evidence of a dominant stricture. Typical cholangiographic findings in PSC are multifocal stricturing and beading involving both the intrahepatic and extrahepatic biliary tree. These strictures are typically short and with intervening segments of normal or slightly dilated bile ducts that are distributed diffusely, producing the classic beaded appearance. The gallbladder and cystic duct can be involved in up to 15% of cases. Gradually, biliary cirrhosis develops, and patients will progress to decompensated liver disease with all the manifestations of ascites, esophageal variceal hemorrhage, and encephalopathy.

TREATMENT

Primary Sclerosing Cholangitis

There is no specific proven treatment for PSC. Some clinicians use UDCA at “PBC dosages” of 13–15 mg/kg per d with anecdotal improvement, although no study has shown convincing evidence of clinical benefit. A study of high-dose (28–30 mg/kg per d) UDCA found it to be harmful. Endoscopic dilatation of dominant strictures can be helpful, but the ultimate treatment is liver transplantation when decompensated cirrhosis develops. Episodes of cholangitis should be treated with antibiotics. A dreaded complication of PSC is the development of cholangiocarcinoma, which is a relative contraindication to liver transplantation.

CARDIAC CIRRHOSIS

Definition Patients with long-standing right-sided congestive heart failure may develop chronic liver injury and cardiac cirrhosis. This is an increasingly uncommon, if not rare, cause of chronic liver disease given the advances made in the care of patients with heart failure.

Etiology and Pathology In the case of long-term right-sided heart failure, there is an elevated venous pressure transmitted via the inferior vena cava and hepatic veins to the sinusoids of the liver, which become dilated and engorged with blood. The liver becomes enlarged and swollen, and with long-term passive congestion and relative ischemia due to poor circulation, centrilobular hepatocytes can become necrotic, leading to pericentral fibrosis. This fibrotic pattern can extend to the periphery of the lobule outward until a unique pattern of fibrosis causing cirrhosis can occur.

Clinical Features Patients typically have signs of congestive heart failure and will manifest an enlarged firm liver on physical examination. ALP levels are characteristically elevated, and aminotransferases may be normal or slightly increased, with AST usually higher than ALT. It is unlikely that patients will develop variceal hemorrhage or encephalopathy.

Diagnosis The diagnosis is usually made in someone with clear-cut cardiac disease who has an elevated ALP and an enlarged liver. Liver biopsy shows a pattern of fibrosis that can be recognized by an experienced hepatopathologist. Differentiation from Budd-Chiari syndrome (BCS) can be made by seeing extravasation of red blood cells in BCS, but not in cardiac hepatopathy. Veno-occlusive disease, now termed sinusoidal obstructive syndrome, can also affect hepatic outflow and has characteristic features on liver biopsy. Sinusoidal obstructive syndrome can be seen under the circumstances of conditioning for bone marrow transplant with radiation and chemotherapy; it can also be seen with the ingestion of certain herbal teas as well as pyrrolizidine alkaloids. This is typically seen in Caribbean countries and rarely in the United States. Treatment is based on management of the underlying cardiac disease.

OTHER TYPES OF CIRRHOSIS

There are several other less common causes of chronic liver disease that can progress to cirrhosis. These include inherited metabolic liver diseases such as hemochromatosis, Wilson’s disease, α_1 antitrypsin (α_1 AT) deficiency, and cystic fibrosis. For these disorders, the manifestations of cirrhosis are similar, with some minor variations, to those seen in other patients with other causes of cirrhosis.

Hemochromatosis is an inherited disorder of iron metabolism that results in a progressive increase in hepatic iron deposition, which, over time, can lead to a portal-based fibrosis progressing to cirrhosis, liver failure, and hepatocellular cancer. While the frequency of hemochromatosis is relatively common, with genetic susceptibility occurring in 1 in 250 individuals, the frequency of end-stage manifestations due to the disease is relatively low, and <5% of those patients who are genetically susceptible will go on to develop severe liver disease from hemochromatosis. Diagnosis is made with serum iron studies showing an elevated transferrin saturation and an elevated ferritin level, along with abnormalities identified by HFE mutation analysis. Treatment is straightforward, with regular therapeutic phlebotomy.

Wilson’s disease is an inherited disorder of copper homeostasis with failure to excrete excess amounts of copper, leading to an accumulation in the liver. This disorder is relatively uncommon, affecting 1 in 30,000 individuals. Wilson’s disease typically affects adolescents and young adults. Prompt diagnosis before end-stage manifestations become irreversible can lead to significant clinical improvement. Diagnosis requires determination of ceruloplasmin levels, which are low; 24-h urine copper levels, which are elevated; typical physical examination findings, including Kayser-Fleischer corneal rings; and characteristic liver biopsy findings. Treatment consists of copper-chelating medications.

α_1 AT deficiency results from an inherited disorder that causes abnormal folding of the α_1 AT protein, resulting in failure of secretion of that

protein from the liver. It is unknown how the retained protein leads to liver disease. Patients with α_1 AT deficiency at greatest risk for developing chronic liver disease have the ZZ phenotype, but only ~10–20% of such individuals will develop chronic liver disease. Diagnosis is made by determining α_1 AT levels and phenotype. Characteristic periodic acid–Schiff (PAS)–positive, diastase-resistant globules are seen on liver biopsy. The only effective treatment is liver transplantation, which is curative.

Cystic fibrosis is an uncommon inherited disorder affecting whites of northern European descent. A biliary-type cirrhosis can occur, and some patients derive benefit from the chronic use of UDCA.

MAJOR COMPLICATIONS OF CIRRHOsis

These include gastroesophageal variceal hemorrhage, splenomegaly, ascites, hepatic encephalopathy, spontaneous bacterial peritonitis (SBP), hepatorenal syndrome (HRS), and hepatocellular carcinoma (Table 344-2). There are also more rare complications in the pulmonary system including hepatopulmonary syndrome and portopulmonary hypertension.

■ PORTAL HYPERTENSION

Portal hypertension is defined as the elevation of the hepatic venous pressure gradient (HVPG) to >5 mmHg. Portal hypertension is caused by a combination of two simultaneously occurring hemodynamic processes: (1) increased intrahepatic resistance to the passage of blood flow through the liver due to cirrhosis, regenerative nodules, and microthrombi, and (2) increased splanchnic blood flow secondary to vasodilation within the splanchnic vascular bed. In more advanced stages, there is also activation of neurohumoral responses and vasoconstrictive systems resulting in sodium and water retention, increased blood volume, and hyperdynamic circulatory system producing more portal hypertension. There is usually an initial stage of compensated cirrhosis with HVPG between 5 and 10 mmHg that can be asymptomatic and last for ≥10 years, but when clinically significant portal hypertension develops (defined as a HVPG ≥10 mmHg), there is substantial risk of decompensation with variceal bleeding, ascites, or hepatic encephalopathy. With decompensation, median mortality is <2 years. *Variceal hemorrhage* is an immediate life-threatening problem with a 20–30% mortality rate associated with each episode of bleeding. The portal venous system normally drains blood from most of the GI tract including the stomach, small and large intestines, spleen, pancreas, and gallbladder.

The causes of portal hypertension are usually subcategorized as prehepatic, intrahepatic, and posthepatic (Table 344-3). Prehepatic causes of portal hypertension are those affecting the portal venous system before it enters the liver; they include portal vein thrombosis and splenic vein thrombosis. Posthepatic causes encompass those affecting the hepatic veins and venous drainage to the heart; they include BCS and chronic right-sided cardiac congestion. Intrahepatic causes account for >95% of cases of portal hypertension and are represented by the major forms of cirrhosis. Intrahepatic causes of portal hypertension can be further subdivided into presinusoidal, sinusoidal, and postsinusoidal causes. Postsinusoidal causes include veno-occlusive

TABLE 344-3 Classification of Portal Hypertension

Prehepatic	
	Portal vein thrombosis
	Splenic vein thrombosis
	Massive splenomegaly (Banti's syndrome)
Hepatic	
	Presinusoidal
	Schistosomiasis
	Congenital hepatic fibrosis
	Sinusoidal
	Cirrhosis—many causes
	Alcoholic hepatitis
	Postsinusoidal
	Hepatic sinusoidal obstruction (veno-occlusive syndrome)
Posthepatic	
	Budd-Chiari syndrome
	Inferior vena caval webs
	Cardiac causes
	Restrictive cardiomyopathy
	Constrictive pericarditis
	Severe congestive heart failure

disease, whereas presinusoidal causes include congenital hepatic fibrosis and schistosomiasis. Sinusoidal causes are related to cirrhosis from various causes.

Cirrhosis is the most common cause of portal hypertension in the United States, and clinically significant portal hypertension is present in >60% of patients with cirrhosis. Portal vein obstruction may be idiopathic or can occur in association with cirrhosis or with infection, pancreatitis, or abdominal trauma.

Coagulation disorders that can lead to the development of portal vein thrombosis include polycythemia vera; essential thrombocythemia; deficiencies in protein C, protein S, antithrombin III, and factor V Leiden; and abnormalities in the gene-regulating prothrombin production. Some patients may have a subclinical myeloproliferative disorder.

Clinical Features The three primary complications of portal hypertension are gastroesophageal varices with hemorrhage, ascites, and hypersplenism. Thus, patients may present with upper GI bleeding, which, on endoscopy, is found to be due to esophageal or gastric varices; with the development of ascites along with peripheral edema; or with an enlarged spleen with associated reduction in platelets and white blood cells on routine laboratory testing.

ESOPHAGEAL VARICES Over the past decade, it has become common practice to screen known cirrhotics with endoscopy to look for esophageal varices. Such screening studies have shown that approximately one-third of patients with histologically confirmed cirrhosis have varices. Approximately 5–15% of cirrhotics per year develop varices, and it is estimated that the majority of patients with cirrhosis will develop varices over their lifetimes. Furthermore, it is anticipated that roughly one-third of patients with varices will develop bleeding. Several factors predict the risk of bleeding, including the severity of cirrhosis (Child-Pugh class, Model for End-Stage Liver Disease [MELD] score); the height of wedged-hepatic vein pressure; the size of the varix; the location of the varix; and certain endoscopic stigmata, including red wale signs, hematocystic spots, diffuse erythema, bluish color, cherry red spots, or white-nipple spots. Patients with tense ascites are also at increased risk for bleeding from varices.

Diagnosis In patients with cirrhosis who are being followed chronically, the development of portal hypertension is usually revealed by the presence of thrombocytopenia; the appearance of an enlarged spleen; or the development of ascites, encephalopathy, and/or esophageal varices with or without bleeding. In previously undiagnosed patients, any of these features should prompt further evaluation to determine the presence of portal hypertension and liver disease. Varices should

TABLE 344-2 Complications of Cirrhosis

Portal hypertension	Coagulopathy
Gastroesophageal varices	Factor deficiency
Portal hypertensive gastropathy	Fibrinolysis
Splenomegaly, hypersplenism	Thrombocytopenia
Ascites	Bone disease
Spontaneous bacterial peritonitis	Osteopenia
Hepatorenal syndrome	Osteoporosis
Type 1	Osteomalacia
Type 2	Hematologic abnormalities
Hepatic encephalopathy	Anemia
Hepatopulmonary syndrome	Hemolysis
Portopulmonary hypertension	Thrombocytopenia
Malnutrition	Neutropenia

2630 be identified by endoscopy. Contrasted-enhanced abdominal imaging, either by computed tomography (CT) or MRI, can be helpful in demonstrating a nodular liver and in finding changes of portal hypertension with intraabdominal collateral circulation. Rarely, the HVPG is measured by interventional radiology. Patients with a gradient >12 mmHg are at risk for variceal hemorrhage.

TREATMENT

Variceal Hemorrhage

Treatment for esophageal varices as a complication of portal hypertension is divided into two main categories: (1) primary prophylaxis and (2) prevention of rebleeding once there has been an initial variceal hemorrhage. Primary prophylaxis requires routine surveillance by endoscopy of all patients with cirrhosis. Upper endoscopies are recommended at diagnosis of compensated cirrhosis and then every 2 years if the liver disease is active or every 3 years if inactive (alcohol cessation, viral hepatitis eradication). Endoscopy is also recommended at the time of hepatic decompensation. Once varices that are at increased risk for bleeding are identified, usually defined as medium or large varices or small varices with high-risk stigmata or in decompensated cirrhosis, primary prophylaxis can be achieved either through nonselective beta blockade (NSBB) titrated with a goal heart rate of 55–60 beats/min with systolic blood pressure >90 mmHg or by variceal band ligation. Numerous placebo-controlled clinical trials of either propranolol or nadolol show a lower risk of variceal hemorrhage and mortality related to variceal hemorrhage but no clear benefit on overall survival.

Endoscopic variceal ligation (EVL) has been compared to NSBB for primary prophylaxis against variceal bleeding, and EVL appears to have equivalent efficacy. Two more recent trials comparing EVL to carvedilol, a drug with NSBB and anti- α_1 -adrenergic properties, showed similar efficacy. Thus, either NSSB or EVL is effective for primary prophylaxis of bleeding, and the choice should be based on patient and physician preference and tolerability. Once primary prophylaxis has been initiated, repeat endoscopy for surveillance of varices is unnecessary.

The approach to patients once they have had a variceal bleed is first to treat the acute bleed, which can be life-threatening, and then to prevent further bleeding. Treatment of acute bleeding requires both fluid and red blood cell replacement to stabilize hemodynamics. A recent randomized trial of restricted transfusion starting when hemoglobin is <7 g/dL with a goal hemoglobin of 7–9 g/dL, compared to a more liberal strategy, resulted in reduced early rebleeding and mortality. This strategy is recommended, although adjustments should be made based on cardiac risks and hemodynamics. Correcting an elevated prothrombin time with fresh frozen plasma is not recommended unless there is evidence of coagulopathy (bleeding at other sites such as IV lines). The use of vasoconstricting agents, usually somatostatin or octreotide, has been shown to improve initial bleeding control and reduce transfusion requirements and all-cause mortality. Prophylactic antibiotics, usually with ceftriaxone, started prior to endoscopy result in reduced infections, recurrent bleeding, and mortality. Balloon tamponade (Sengstaken-Blakemore tube or Minnesota tube) can be used in patients who need stabilization prior to endoscopic therapy or as a bridge to transjugular intrahepatic portosystemic shunt (TIPS) after endoscopic failure. Control of bleeding can be achieved in the vast majority of cases; however, bleeding recurs in the majority of patients if definitive endoscopic therapy has not been instituted. Upper endoscopy is used as first-line treatment to diagnose the cause of the bleeding and to control bleeding acutely with EVL. When esophageal varices extend into the proximal stomach or the bleeding varices are entirely within the stomach, band ligation is often unsuccessful. In these situations, consideration for a TIPS should be made. This technique creates a portosystemic shunt by a percutaneous approach using an expandable metal stent, which is advanced under angiographic guidance to the hepatic veins and then through the substance of the liver to create a direct portacaval

shunt. Encephalopathy can occur in as many as 20% of patients after TIPS and is particularly problematic in elderly patients and in patients with preexisting encephalopathy. TIPS is usually reserved for individuals who fail or are unable to receive endoscopic therapy, although there is emerging evidence that patient who are highly selected to be at high risk for rebleeding may also benefit. TIPS can sometimes be used as a bridge to transplantation, and all patients requiring TIPS should be considered for transplant evaluation. Some gastric varices are associated with a splenorenal shunt and can be effectively treated with a balloon occluded retrograde transvenous obliteration (BRTO) of varices sometimes in combination with a TIPS. Prevention of further bleeding is usually accomplished with repeated variceal band ligation until varices are obliterated in combination with NSBB. If recurrent variceal bleeding occurs, then TIPS should be performed for long-term prevention of bleeding. Once a TIPS has been performed, there is no need for further endoscopies for variceal surveillance; however, the TIPS should be periodically monitored with Doppler ultrasound for stenosis. (Fig. 344-3).

PORTAL HYPERTENSIVE GASTROPATHY

Portal hypertensive gastropathy can cause both acute clinical GI bleeding and chronic bleeding resulting in iron-deficiency anemia. It is associated with all causes of portal hypertension and is diagnosed by characteristic endoscopy findings showing a snake skin-like mosaic pattern of gastric mucosa often with central red or brown spots. When there is bleeding, treatment is with NSBB and iron repletion. Refractory bleeding may respond to TIPS.

SPLENOMEGALY AND HYPERSPLENISM

Congestive splenomegaly with hypersplenism is common in patients with portal hypertension and is usually the first indication of portal hypertension. Clinical features include the presence of an enlarged spleen on physical examination and the development of thrombocytopenia and leukopenia in patients who have cirrhosis. Some patients will have significant left-sided and left upper quadrant abdominal pain

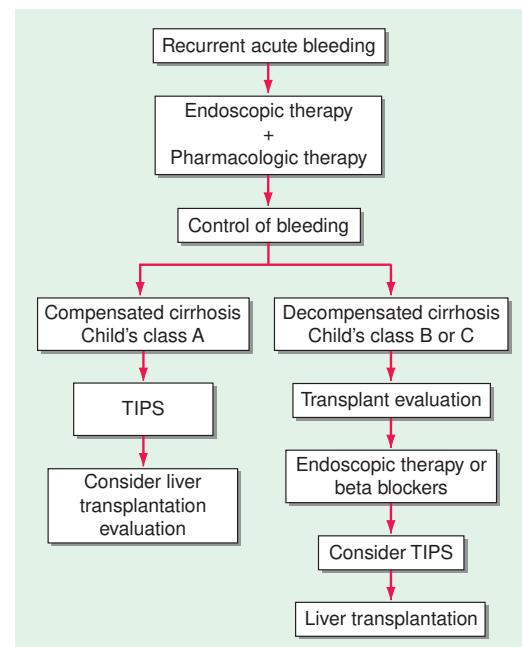


FIGURE 344-3 Management of recurrent variceal hemorrhage. This algorithm describes an approach to management of patients who have recurrent bleeding from esophageal varices. Initial therapy is generally with endoscopic therapy often supplemented by pharmacologic therapy. With control of bleeding, a decision needs to be made as to whether patients should go on to transjugular intrahepatic portosystemic shunt (TIPS; if they are Child's class A) or if they should have TIPS and be considered for transplant (if they are Child's class B or C).

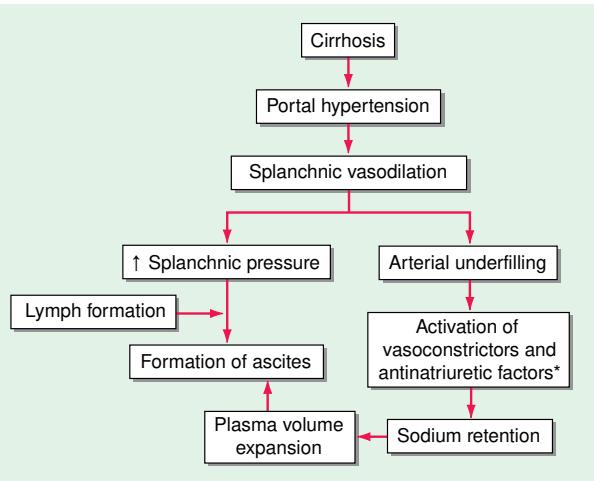


FIGURE 344-4 Development of ascites in cirrhosis. This flow diagram illustrates the importance of portal hypertension with splanchnic vasodilation in the development of ascites. *Antinatriuretic factors include the renin-angiotensin-aldosterone system and the sympathetic nervous system.

related to an enlarged spleen. Splenomegaly itself usually requires no specific treatment.

ASCITES

Definition Ascites is the accumulation of fluid within the peritoneal cavity. Overwhelmingly, the most common cause of ascites is portal hypertension related to cirrhosis; however, clinicians should remember that malignant, infectious, and cardiac causes of ascites can be present as well, and careful differentiation of these other causes is obviously important for patient care.

Pathogenesis The presence of portal hypertension contributes to the development of ascites in patients who have cirrhosis (Fig. 344-4). There is an increase in intrahepatic resistance, causing increased portal pressure, but there is also vasodilation of the splanchnic arterial system, which, in turn, results in an increase in portal venous inflow. Both abnormalities result in increased production of splanchnic lymph. Vasodilating factors such as nitric oxide are responsible for the vasodilatory effect. There is activation of the renin-angiotensin-aldosterone system with the development of hyperaldosteronism and activation of the sympathetic nervous system as a consequence of a homeostatic response caused by underfilling of the arterial circulation secondary to arterial vasodilation in the splanchnic vascular bed. The renal effects of increased aldosterone and activation of the sympathetic nervous system lead to sodium retention causing fluid accumulation and expansion of the extracellular fluid volume, resulting in peripheral edema and ascites. Because the retained fluid is constantly leaking out of the intravascular compartment into the peritoneal cavity, the sensation of vascular filling is not achieved, and the process continues. Hypoalbuminemia from decreased synthetic function in a cirrhotic liver results in reduced plasma oncotic pressure and contributes to the loss of fluid from the vascular compartment into the peritoneal cavity.

Clinical Features Patients typically note an increase in abdominal girth that is often accompanied by the development of peripheral edema. The development of ascites is often insidious, and it is surprising that some patients wait so long and become so distended before seeking medical attention. Patients usually have at least 1–2 L of fluid in the abdomen before they are aware that there is an increase. If ascitic fluid is massive, respiratory function can be compromised, causing dyspnea. Hepatic hydrothorax may also contribute to respiratory symptoms. Patients with massive ascites are often malnourished and have muscle wasting and excessive fatigue and weakness.

Diagnosis Diagnosis of ascites is by physical examination and is often aided by abdominal imaging. Patients will have bulging flanks,

may have a fluid wave, or may have the presence of shifting dullness. This is determined by taking patients from a supine position to lying on either their left or right side and noting the movement of the dullness to percussion. Subtle amounts of ascites can be detected by ultrasound or CT scanning. Hepatic hydrothorax is more common on the right side and implicates a rent in the diaphragm with free flow of ascitic fluid into the thoracic cavity.

When patients present with ascites for the first time, it is recommended that a diagnostic paracentesis be performed to characterize the fluid. This should include the determination of total protein and albumin content, blood cell counts with differential, and cultures. In the appropriate setting, amylase may be measured and cytology performed. In patients with cirrhosis, the protein concentration of the ascitic fluid is low, usually <2.5 g/dL. The serum ascites-to-albumin gradient (SAAG), calculated by subtracting the fluid albumin level from the serum albumin level, has replaced the description of exudative or transudative fluid. When the SAAG is >1.1 g/dL, the cause of the ascites is most likely due to portal hypertension; this is usually in the setting of cirrhosis. Cardiac ascites can be identified by SAAG >1.1 g/dL and ascites protein >2.5 g/dL. When the SAAG is <1.1 g/dL, infectious or malignant causes of ascites should be considered. When ascitic fluid protein is very low, <1.5 g/dL, patients are at increased risk for developing SBP. A high level of red blood cells in the ascitic fluid usually signifies a traumatic tap but can also rarely occur with hepatocellular cancer or a ruptured omental varix. When the absolute level of polymorphonuclear leukocytes is >250/ μ L, infection is likely.

TREATMENT

Ascites

Patients with small amounts of ascites can usually be managed with dietary sodium restriction alone. Most average diets in the United States contain 6–8 g of sodium per day, and if patients eat at restaurants or fast-food outlets, the amount of sodium in their diet can exceed this amount. Thus, it is often extremely difficult to get patients to change their dietary habits to ingest 2 g of sodium per day, equivalent to slightly more than three-quarters of a teaspoon of salt, which is the recommended amount. Sodium educational pamphlets are helpful. Often, a simple recommendation is to eat fresh or frozen foods, avoiding canned or processed foods. When a moderate amount of ascites is present, diuretic therapy is usually necessary. Traditionally, spironolactone at 100 mg/d as a single dose is started, and furosemide may be added at 40 mg/d, particularly in patients who have peripheral edema. Failure of the diuretics suggests that patients may not be compliant with a low-sodium diet. If compliance is confirmed and ascitic fluid is not being mobilized, there should be incremental increases in spironolactone to a maximum of 400 mg/d and furosemide to 160 mg/d. If a large amount of ascites is still present on diuretics in patients who are compliant with a low-sodium diet, then they are defined as having *refractory ascites*, and alternative treatment modalities including repeated large-volume paracentesis (LVP) or a TIPS procedure should be considered (Fig. 344-5). After LVP of ≥ 5 L, IV 25% albumin at a dose of ~8 g/L of removed ascites should be given to prevent circulatory dysfunction. Multiple studies have shown that TIPS, although effective at managing the ascites, does not improve survival. Unfortunately, TIPS is often associated with an increased frequency of hepatic encephalopathy and must be considered carefully on a case-by-case basis. The prognosis for patients with cirrhosis with ascites is poor, and some studies have shown that <50% of patients survive 2 years after the onset of ascites. Thus, there should be consideration for liver transplantation in patients with ascites. Patients with cirrhosis and ascites are at increased risk for renal failure from certain medications including nonsteroidal anti-inflammatory drugs and aminoglycosides; therefore, these medications should generally be avoided. Angiotensin-converting enzyme inhibitors and angiotensin receptor blockers should be used cautiously with close monitoring of blood pressure and renal function.

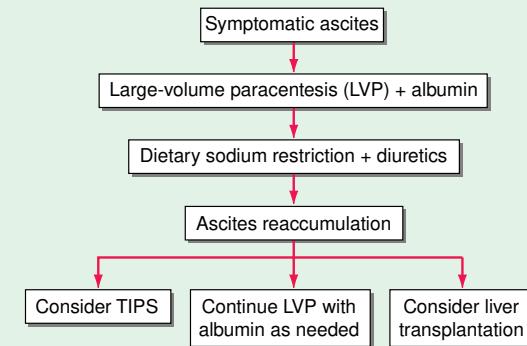


FIGURE 344-5 Treatment of refractory ascites. In patients who develop azotemia in the course of receiving diuretics in the management of their ascites, some will require repeated large-volume paracentesis (LVP), some may be considered for transjugular intrahepatic portosystemic shunt (TIPS), and some would be good candidates for liver transplantation. These decisions are all individualized.

■ SPONTANEOUS BACTERIAL PERITONITIS

SBP is a common and severe complication of ascites characterized by spontaneous infection of the ascitic fluid without an intraabdominal source. In hospitalized patients with cirrhosis and ascites, SBP can occur in up to 30% of individuals and can have a 25% in-hospital mortality rate. Bacterial translocation is the presumed mechanism for development of SBP, with gut flora traversing the intestine into mesenteric lymph nodes, leading to bacteremia and seeding of the ascitic fluid. The most common organisms are *Escherichia coli* and other gut bacteria; however, gram-positive bacteria, including *Streptococcus viridans*, *Staphylococcus aureus*, and *Enterococcus* spp., can also be found. If more than two organisms are identified, secondary bacterial peritonitis due to a perforated viscus should be considered. The diagnosis of SBP is made when the fluid sample has an absolute neutrophil count >250/ μ L. Bedside cultures should be obtained by direct injection of ascitic fluid into blood culture bottles. Patients with ascites may present with fever, altered mental status, elevated white blood cell count, abdominal pain or discomfort, and acute kidney injury, or they may present without any of these features. Therefore, it is necessary to have a high degree of clinical suspicion, and peritoneal taps are recommended for most cirrhosis patients hospitalized with ascites and cirrhosis complications or signs of infection. Treatment is commonly with intravenous third-generation cephalosporin for 5 days. In addition, intravenous albumin (1.5 g/kg body weight on day and 1.0 g/kg on day 3) has been shown to reduce the risk of renal failure and to improve survival. In patients with variceal hemorrhage, the frequency of SBP is significantly increased, and prophylaxis against SBP is recommended when a patient presents with upper GI bleeding. Furthermore, in patients who have had an episode (or multiple episodes) of SBP and recovered, quinolone antibiotic prophylaxis should be given to prevent recurrent SBP.

■ HEPATORENAL SYNDROME

HRS is a form of functional renal failure without renal pathology that occurs in ~10% of patients with advanced cirrhosis or acute liver failure. There are marked disturbances in the arterial renal circulation in patients with HRS; these include an increase in vascular resistance accompanied by a reduction in systemic vascular resistance. The reason for renal vasoconstriction is most likely multifactorial and is poorly understood. The diagnosis is made usually in the presence of a large amount of ascites in patients who have a stepwise progressive increase in creatinine. Type 1 HRS is characterized by a progressive impairment in renal function and a significant reduction in creatinine clearance within 1–2 weeks of presentation. Type 2 HRS is characterized by a reduction in glomerular filtration rate with an elevation of serum creatinine level, but it is stable and is associated with a better outcome than that of type 1 HRS.

HRS requires exclusion of other causes of acute renal failure, most notably volume depletion. Diuretics should be stopped, and infusion of albumin 1 g/kg per day is recommended. Treatment is with vasoconstrictors such as terlipressin (not currently available in North America) or low-dose norepinephrine (requires intensive care unit monitoring). Midodrine, an α -agonist, along with octreotide and intravenous albumin are also commonly used in the United States. The best therapy for HRS is liver transplantation; recovery of renal function is typical in this setting. In patients with either type 1 or type 2 HRS, the prognosis is poor unless transplant can be achieved within a short period of time.

■ HEPATIC ENCEPHALOPATHY

Portosystemic encephalopathy is a serious complication of chronic liver disease and is broadly defined as an alteration in mental status and cognitive function occurring in the presence of liver failure. In severe acute liver injury, the development of encephalopathy is a requirement for a diagnosis of acute liver failure and can be seen in association with life-threatening brain edema, which is not a feature in chronic liver disease. Encephalopathy is much more commonly seen in patients with chronic liver disease. Gut-derived neurotoxins that are not removed by the liver because of vascular shunting and decreased hepatic mass reach the brain and cause the symptoms known as hepatic encephalopathy. Ammonia levels are typically elevated, but the correlation between severity of liver disease and height of ammonia levels is often poor, and most hepatologists do not rely on ammonia levels to make a diagnosis or follow clinical progress. Other compounds and metabolites that may contribute to the development of encephalopathy include certain false neurotransmitters and mercaptans.

Clinical Features In acute liver failure, changes in mental status can occur rapidly. Brain edema can be seen in these patients, with severe encephalopathy associated with swelling of the gray matter. Cerebral herniation is a feared complication of brain edema in acute liver failure, and treatment to decrease edema is with mannitol and judicious use of intravenous fluids.

In patients with cirrhosis, encephalopathy is often found as a result of certain precipitating events such as hypokalemia, infection, an increased dietary protein load, or volume depletion. Patients may be confused or exhibit a change in personality. They may actually be quite violent and difficult to manage; alternatively, patients may be very sleepy and difficult to rouse. Precipitating events are common, so they should be sought carefully. If patients have ascites, this should be tapped to rule out infection. Evidence of GI bleeding should be sought, and patients should be appropriately hydrated. Electrolytes should be measured and abnormalities corrected. In patients presenting with encephalopathy, asterixis is often present. Asterixis can be elicited by having patients extend their arms and bend their wrists back. Patients who are encephalopathic have a “liver flap”—that is, a sudden forward movement of the wrist. This requires patients to be able to cooperate with the examiner. Alternative causes for altered mental status should also be considered.

The diagnosis of hepatic encephalopathy is clinical and requires an experienced clinician to recognize and put together all the various features. Often when patients have encephalopathy for the first time, they (and/or their caregivers) are unaware of what is transpiring, but once they have been through the experience, they can identify when this is developing in subsequent situations and can often self-medicate to prevent the development or worsening of encephalopathy.

TREATMENT

Hepatic Encephalopathy

Treatment is multifactorial and includes management of the above-mentioned precipitating factors. Sometimes hydration and correction of electrolyte imbalance are all that is necessary. In the past, restriction of dietary protein was used; however, the negative impact of that maneuver on overall nutrition is thought to outweigh the benefit, and it is thus strongly discouraged. The mainstay of

treatment for encephalopathy is to use lactulose, a nonabsorbable disaccharide, which results in colonic acidification. Catharsis ensues, contributing to the elimination of nitrogenous products in the gut that are responsible for the development of encephalopathy. The goal of lactulose therapy is to promote two to three soft stools per day. Patients are asked to titrate their amount of ingested lactulose to achieve the desired effect. Lactulose is usually continued after the initial episode of encephalopathy. Poorly absorbed antibiotics are often used as adjunctive therapies for patients who have a difficult time with lactulose. The alternating administration of neomycin and metronidazole has been used in the past to reduce the individual side effects of each: neomycin for renal insufficiency and ototoxicity and metronidazole for peripheral neuropathy. More recently, rifaximin at 550 mg twice daily has been very effective in preventing recurrent encephalopathy. Zinc supplementation is sometimes helpful and is relatively harmless. The development of encephalopathy in patients with chronic liver disease is a poor prognostic sign, but this complication can be managed in the vast majority of patients.

LIVER LUNG SYNDROMES

Hepatopulmonary syndrome (HPS) is characterized by arterial hypoxemia in a patient with cirrhosis without significant lung disease. The liver disease causes intrapulmonary vascular dilations resulting in blood shunting past alveoli and significant ventilation-perfusion mismatch. Clinical symptoms include dyspnea and platypnea. HPS is common, occurring in 4–32% of patients with cirrhosis; however, it is often mild. Diagnosis involves demonstrating hypoxemia, without evidence of significant lung disease, and shunt on bubble echocardiography. Treatment is with oxygen supplementation and liver transplantation.

Portopulmonary hypertension (PPHT) is pulmonary hypertension in a patient with portal hypertension. The portal hypertension results in the production of vasoconstrictor substances that affect the pulmonary artery. Many patients are asymptomatic, especially early in the disease; however, they later can develop dyspnea on exertion and fatigue. PPHT is rare, occurring in ~5% of patients with advanced cirrhosis. Diagnosis includes initial identification on echocardiogram and confirmation on right heart catheterization showing elevated mean pulmonary artery pressure, elevated pulmonary vascular resistance, and normal pulmonary capillary wedge pressure. Prognosis is poor, although liver transplantation after effective reduction in pulmonary artery pressure with vasodilatory medications can be effective.

MALNUTRITION IN CIRRHOSES

Because the liver is principally involved in the regulation of protein and energy metabolism in the body, it is not surprising that patients with advanced liver disease are commonly malnourished. Once patients become cirrhotic, they are more catabolic, and muscle protein is metabolized. There are multiple factors that contribute to the malnutrition of cirrhosis, including poor dietary intake, alterations in gut nutrient absorption, and alterations in protein metabolism. Close attention to food intake is helpful in preventing patients from becoming catabolic. General recommendations include multiple small meals including a late evening snack with total calories of 25–30 kcal per kg of ideal body weight per day and 1.2–1.5 g of protein per kg of ideal body weight per day.

ABNORMALITIES IN COAGULATION

Coagulation disorders in cirrhosis are poorly understood, and typical clinical measures of coagulation, such as the prothrombin time and international normalized ratio, are not reliable measures of clotting ability. There is decreased synthesis of both pro- and anticoagulant factors and thus some rebalancing in coagulation; however, the coagulation cascade can easily tip toward thrombosis or bleeding. In addition, patients may have thrombocytopenia from hypersplenism due to portal hypertension and some platelet dysfunction, which is counterbalanced with increased von Willebrand factor. Adequate thrombin formation can occur with platelet levels from cirrhosis patients >50,000–60,000/L. Synthesis of vitamin K-dependent clotting factors II, VII, IX, and X

is diminished in patients with chronic cholestatic syndromes because absorption of vitamin K requires good bile flow. Intravenous or intramuscular vitamin K can quickly correct this abnormality. Overall, the status of coagulation in a cirrhotic patient needs to be judged clinically rather than relying on current laboratory tests.

BONE DISEASE IN CIRRHOSES

Osteoporosis is common in patients with chronic cholestatic liver disease because of malabsorption of vitamin D and decreased calcium ingestion. The rate of bone resorption exceeds that of new bone formation in patients with cirrhosis, resulting in bone loss. Dual-energy x-ray absorptiometry (DEXA) is a useful method for determining osteoporosis or osteopenia. When a DEXA scan shows osteoporosis, treatment with bisphosphonates is effective.

HEMATOLOGIC ABNORMALITIES IN CIRRHOSES

Numerous hematologic manifestations of cirrhosis are present, including anemia from a variety of causes including hypersplenism, hemolysis, iron deficiency, and perhaps folate deficiency from malnutrition. Macrocytosis is a common abnormality in red blood cell morphology seen in patients with chronic liver disease, and neutropenia may be a result of hypersplenism.

FURTHER READING

- AASLD/IDSA HCV G P : Hepatitis C guidance 2019 update: AASLD-IDSA recommendations for testing, managing, and treating hepatitis C virus infection. *Hepatology* 71:686, 2020.
- B SW et al: Diagnosis, evaluation and management of ascites and hepatorenal syndrome: 2021 Practice guidance by the American Association for the Study of Liver Diseases. *Hepatology* 74:1014, 2021.
- G -T 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.
- T NA et al: Update on prevention, diagnosis and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. *Hepatology* 67:1560, 2018.
- V H et al: Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. *Hepatology* 60:715, 2014.

345

Liver Transplantation

Raymond T. Chung, Jules L. Dienstag



Liver transplantation—the replacement of the native, diseased liver by a normal organ (allograft)—has matured from an experimental procedure reserved for desperately ill patients to an accepted, lifesaving operation applied more optimally in the natural history of end-stage liver disease. The preferred and technically most advanced approach is *orthotopic transplantation*, in which the native organ is removed and the donor organ is inserted in the same anatomic location. Pioneered in the 1960s by Thomas Starzl at the University of Colorado and, later, at the University of Pittsburgh and by Roy Calne in Cambridge, England, liver transplantation is now performed routinely worldwide. Success measured as 1-year survival has improved from ~30% in the 1970s to >90% today. These improved prospects for prolonged survival resulted from refinements in operative technique, improvements in organ procurement and preservation, advances in immunosuppressive therapy, and, perhaps most influentially, more enlightened patient selection and timing. Despite the perioperative morbidity and mortality, the technical

and management challenges of the procedure, and its costs, liver transplantation has become the approach of choice for selected patients whose chronic or acute liver disease is progressive, life-threatening, and unresponsive to medical therapy. Based on the current level of success, the number of liver transplants has continued to grow each year; in 2020, 8906 patients received liver allografts in the United States. Still, the demand for new livers continues to outpace availability; as of 2021, 11,710 patients in the United States were on a waiting list for a donor liver. In response to this drastic shortage of donor organs, many transplantation centers supplement deceased-donor liver transplantation with living-donor transplantation.

INDICATIONS

Potential candidates for liver transplantation are children and adults who, in the absence of contraindications (see below), suffer from severe, irreversible liver disease for which alternative medical or surgical treatments have been exhausted or are unavailable. *Timing of the operation is of critical importance.* Indeed, improved timing and better patient selection are felt to have contributed more to the increased success of liver transplantation in the 1980s and beyond than all the impressive technical and immunologic advances combined. Although the disease should be advanced, and although opportunities for spontaneous or medically induced stabilization or recovery should be allowed, the procedure should be done sufficiently early to give the surgical procedure a fair chance for success. Ideally, transplantation should be considered in patients with end-stage liver disease who are experiencing or have experienced a life-threatening complication of hepatic decompensation or whose quality of life has deteriorated to unacceptable levels. Although patients with well-compensated cirrhosis can survive for many years, many patients with quasi-stable chronic liver disease have much more advanced disease than may be apparent. As discussed below, the better the status of the patient prior to transplantation, the higher will be its anticipated success rate. The decision about *when* to transplant is complex and requires the combined judgment of an experienced team of hepatologists, transplant surgeons, anesthesiologists, and specialists in support services, not to mention the well-informed consent of the patient and the patient's family.

TRANSPLANTATION IN CHILDREN

Indications for transplantation in children are listed in Table 345-1. The most common is *biliary atresia*. *Inherited or genetic disorders of metabolism* associated with liver failure constitute another major

indication for transplantation in children and adolescents. In Crigler-Najjar disease type I and in certain hereditary disorders of the urea cycle and of amino acid or lactate-pyruvate metabolism, transplantation may be the only way to prevent impending deterioration of central nervous system function, despite the fact that the native liver is structurally normal. Combined heart and liver transplantation has yielded dramatic improvement in cardiac function and in cholesterol levels in children with homozygous familial hypercholesterolemia; combined liver and kidney transplantation has been successful in patients with primary hyperoxaluria type I. In hemophiliacs with transfusion-associated hepatitis and liver failure, liver transplantation has been associated with recovery of normal factor VIII synthesis.

TRANSPLANTATION IN ADULTS

Liver transplantation is indicated for end-stage *cirrhosis* of all causes (Table 345-1). In *sclerosing cholangitis* and *Caroli's disease* (multiple cystic dilatations of the intrahepatic biliary tree), recurrent infections and sepsis associated with inflammatory and fibrotic obstruction of the biliary tree may be an indication for transplantation. Because prior biliary surgery complicates and is a relative contraindication for liver transplantation, surgical diversion of the biliary tree has been all but abandoned for patients with sclerosing cholangitis. In patients who undergo transplantation for *hepatic vein thrombosis* (*Budd-Chiari syndrome*), postoperative anticoagulation is essential; underlying myeloproliferative disorders may have to be treated but are not a contraindication to liver transplantation. If a donor organ can be located quickly, before life-threatening complications—including cerebral edema—set in, patients with acute liver failure are candidates for liver transplantation. Currently, alcohol-associated liver disease, chronic hepatitis C, and nonalcoholic fatty liver disease (NAFLD) are the most common indications for liver transplantation, accounting for >40% of all adult candidates who undergo the procedure. Patients with alcohol-associated cirrhosis can be considered as candidates for transplantation if they meet strict criteria for abstinence and reform; however, these criteria still do not prevent recidivism in up to a quarter of cases. In highly selected cases in a limited but growing number of centers, transplantation for severe *acute* alcohol-associated hepatitis has been performed with success; however, because patients with acute alcohol-associated hepatitis are still actively using alcohol and because continued alcohol abuse remains a concern, acute alcohol-associated hepatitis is not a routine indication for liver transplantation. Patients with chronic hepatitis C have early allograft and patient survival comparable to those of other subsets of patients after transplantation; however, reinfection in the donor organ is universal, recurrent hepatitis C had been insidiously progressive, with allograft cirrhosis and failure occurring at a higher frequency beyond 5 years. Fortunately, with the introduction of highly effective direct-acting antiviral (DAA) agents targeting hepatitis C virus (HCV), allograft outcomes have improved substantially. In patients with chronic hepatitis B, in the absence of measures to prevent recurrent hepatitis B, survival after transplantation is reduced by ~10–20%; however, prophylactic use of hepatitis B immune globulin (HBIG) during and after transplantation increases the success of transplantation to a level comparable to that seen in patients with nonviral causes of liver decompensation. Specific oral antiviral drugs (e.g., entecavir, tenofovir disoproxil fumarate, tenofovir alafenamide) (Chap. 341) can be used both for prophylaxis against and for treatment of recurrent hepatitis B, facilitating further the management of patients undergoing liver transplantation for end-stage hepatitis B; most transplantation centers rely on antiviral drugs with or without HBIG to manage patients with hepatitis B. Issues of disease recurrence are discussed in more detail below. Patients with nonmetastatic primary hepatobiliary tumors—primary hepatocellular carcinoma (HCC), cholangiocarcinoma, hepatoblastoma, angiosarcoma, epithelioid hemangioendothelioma, and multiple or massive hepatic adenomas—have undergone liver transplantation; however, for some hepatobiliary malignancies, overall survival is significantly lower than that for other categories of liver disease. Most transplantation centers have reported 5-year recurrence-free survival rates in patients with unresectable HCC for single tumors <5 cm in diameter

TABLE 345-1 Indications for Liver Transplantation

CHILDREN	ADULTS
Biliary atresia	Primary biliary cholangitis
Neonatal hepatitis	Secondary biliary cirrhosis
Congenital hepatic fibrosis	Primary sclerosing cholangitis
Alagille's syndrome ^a	Autoimmune hepatitis
Byler's disease ^b	Caroli's disease ^c
α ₁ -Antitrypsin deficiency	Cryptogenic cirrhosis
Inherited disorders of metabolism	Chronic hepatitis with cirrhosis
Wilson's disease	Hepatic vein thrombosis
Tyrosinemia	Fulminant hepatitis
Glycogen storage diseases	Alcohol-associated cirrhosis
Lysosomal storage diseases	Chronic viral hepatitis
Protoporphyrina	Primary hepatocellular malignancies
Crigler-Najjar disease type I	Hepatic adenomas
Familial hypercholesterolemia	Nonalcoholic steatohepatitis
Primary hyperoxaluria type I	Familial amyloid polyneuropathy
Hemophilia	

^aArteriohepatic dysplasia, with paucity of bile ducts, and congenital malformations, including pulmonary stenosis. ^bIntrahepatic cholestasis, progressive liver failure, and mental and growth retardation. ^cMultiple cystic dilatations of the intrahepatic biliary tree.

or for three or fewer lesions all <3 cm comparable to those seen in patients undergoing transplantation for nonmalignant indications. Consequently, liver transplantation is currently restricted to patients whose hepatic malignancies meet these criteria. Expanded criteria for patients with HCC continue to be evaluated. Because the likelihood of recurrent cholangiocarcinoma is very high, only highly selected patients with limited disease are being evaluated for transplantation after intensive chemotherapy and radiation.

CONTRAINDICATIONS

Absolute contraindications for transplantation include life-threatening systemic diseases, uncontrolled extrahepatic bacterial or fungal infections, preexisting advanced cardiovascular or pulmonary disease, multiple uncorrectable life-threatening congenital anomalies, metastatic malignancy, and active drug or alcohol abuse (Table 345-2). Because carefully selected patients in their sixties and even seventies have undergone transplantation successfully, advanced age per se is no longer considered an absolute contraindication; however, in older patients, a more thorough preoperative evaluation should be undertaken to exclude ischemic cardiac disease and other comorbid conditions. Advanced age (>70 years), however, should be considered a *relative contraindication*—that is, a factor to be considered with other relative contraindications. Other relative contraindications include portal vein thrombosis, preexisting renal disease not associated with liver disease (which may prompt consideration of combined liver and kidney transplantation), intrahepatic or biliary sepsis, severe hypoxemia ($P_2 <50$ mmHg) resulting from right-to-left intrapulmonary shunts, portopulmonary hypertension with high mean pulmonary artery pressures (>35 mmHg), previous extensive hepatobiliary surgery, any uncontrolled serious psychiatric disorder, and lack of sufficient social supports. Any one of these relative contraindications is insufficient in and of itself to preclude transplantation. For example, the problem of portal vein thrombosis can be overcome by constructing a graft from the donor liver portal vein to the recipient's superior mesenteric vein. Now that combination antiretroviral therapy has dramatically improved the survival of persons with HIV infection (Chap. 202), and because end-stage liver disease caused by chronic hepatitis C and B has emerged as a serious source of morbidity and mortality in the HIV-infected population, liver transplantation has now been performed successfully

in selected HIV-positive persons who have excellent control of HIV infection. Selected patients with CD4+ T-cell counts >100/ μ L and with pharmacologic suppression of HIV viremia have undergone transplantation for end-stage liver disease. HIV-infected persons who have received liver allografts for end-stage liver disease resulting from chronic hepatitis B have experienced survival rates comparable to those of HIV-negative persons undergoing transplantation for the same indication. In contrast, recurrent HCV in the allograft has until recently limited long-term success in persons with HCV-related end-stage liver disease. Again, the availability of DAA agents targeting HCV (see below and Chap. 341) is expected to improve allograft outcomes significantly.

TECHNICAL CONSIDERATIONS

■ DECEASED DONOR SELECTION

Deceased-donor livers for transplantation are procured primarily from victims of head trauma. Organs from brain-dead donors up to age 60 are acceptable if the following criteria are met: hemodynamic stability, adequate oxygenation, absence of bacterial or fungal infection, absence of abdominal trauma, absence of hepatic dysfunction, and serologic exclusion of hepatitis B virus (HBV), HCV, and HIV. Occasionally, organs from donors with hepatitis B and C are used, particularly for recipients with prior hepatitis B and C, respectively. Organs from donors with antibodies to hepatitis B core antigen (anti-HBc) can also be used when the need is especially urgent, and recipients of these organs are treated prophylactically with antiviral drugs. Increasingly, with the early administration of highly effective DAA agents, organs from donors with hepatitis C have been used successfully in previously uninfected recipients. Cardiovascular and respiratory functions are maintained artificially until the liver can be removed. Transplantation of organs procured from deceased donors who have succumbed to cardiac death can be performed successfully under selected circumstances, when ischemic time is minimized and liver histology preserved. Encouraging improvements in normothermic ex vivo liver perfusion techniques may make broader use of these organs possible. Compatibility in ABO blood group and organ size between donor and recipient are important considerations in donor selection; however, ABO-incompatible, split-liver, or reduced-donor-organ allografts can be performed in emergencies or marked donor scarcity. Tissue typing for human leukocyte antigen (HLA) matching is not required, and preformed cytotoxic HLA antibodies do not preclude liver transplantation. Following perfusion with cold electrolyte solution, the donor liver is removed and packed in ice. The use of University of Wisconsin (UW) solution, rich in lactobionate and raffinose, has permitted the extension of cold ischemic time up to 20 h; however, 12 h may be a more reasonable limit. Improved techniques for harvesting multiple organs from the same donor have increased the availability of donor livers, but the availability of donor livers is far outstripped by the demand. Currently in the United States, all donor livers are distributed through a nationwide organ-sharing network (United Network for Organ Sharing [UNOS]) designed to allocate available organs based on regional considerations and recipient acuity. Recipients who have the highest disease severity generally have the highest priority, but allocation strategies that balance highest urgency against best outcomes continue to evolve to distribute deceased-donor organs most effectively. Allocation based on the Child-Turcotte-Pugh (CTP) score, which uses five clinical variables (encephalopathy stage, ascites, bilirubin, albumin, and prothrombin time) and waiting time, has been replaced by allocation based on urgency alone, calculated using the Model for End-Stage Liver Disease (MELD) score. The MELD score is based on a mathematical model that includes bilirubin, creatinine, and prothrombin time expressed as international normalized ratio (INR) (Table 345-3). Neither waiting time (except as a tie breaker between two potential recipients with the same MELD scores) nor posttransplantation outcome is taken into account, but use of the MELD score has been shown to reduce waiting list mortality, to reduce waiting time prior to transplantation, to be the best predictor of pretransplantation mortality, to satisfy the prevailing view that medical need should be the

TABLE 345-2 Contraindications to Liver Transplantation

ABSOLUTE	RELATIVE
Uncontrolled extrahepatobiliary infection	Age >70
Active, untreated sepsis	Prior extensive hepatobiliary surgery
Uncorrectable, life-limiting congenital anomalies	Portal vein thrombosis
Active substance abuse	Renal failure not attributable to liver disease (consider dual organ transplantation)
Advanced cardiopulmonary disease	Previous extrahepatic malignancy (not including nonmelanoma skin cancer)
Extrahepatobiliary malignancy (not including nonmelanoma malignancy skin cancer)	Severe obesity
Metastatic malignancy to the liver	Severe malnutrition/wasting
Cholangiocarcinoma (except those tumors that fit into protocols)	Medical noncompliance
AIDS	HIV seropositivity with failure to control HIV viremia or CD4 <100/ μ L
Life-threatening systemic diseases	Intrahepatic sepsis Severe hypoxemia secondary to right-to-left intrapulmonary shunts ($P_2 <50$ mmHg)
	Severe pulmonary hypertension (mean pulmonary artery pressure >35 mmHg)
	Uncontrolled psychiatric disorder

**TABLE 345-3 United Network for Organ Sharing (UNOS)
Liver Transplantation Waiting List Criteria**

Status 1	Fulminant hepatic failure (including primary graft nonfunction and hepatic artery thrombosis within 7 days after transplantation as well as acute decompensated Wilson's disease) ^a
----------	--

The Model for End-Stage Liver Disease (MELD)-Na score, on a continuous scale,^b determines allocation of the remainder of donor organs. This model is based on the following calculation:

$$\text{MELD} = 3.78 \times \log_{10} \text{bilirubin (mg/100 mL)} + 11.2 \times \log_{10} \text{international normalized ratio (INR)} + 9.57 \times \log_{10} \text{creatinine (mg/100 mL)} + 6.43^{c,d,e}$$

$$\text{MELD-Na} = \text{MELD} + 1.32 \times (137 - \text{Na [meq/L]}) - [0.033 \times \text{MELD} \times (137 - \text{Na [meq/L]})]$$

Online calculators to determine MELD scores are available, such as the following: <https://optn.transplant.hrsa.gov/resources/allocation-calculators/meld-calculator/>

^aFor children <18 years of age, status 1 includes acute or chronic liver failure plus hospitalization in an intensive care unit or inborn errors of metabolism. Status 1 is retained for those persons with fulminant hepatic failure and supersedes the MELD score. ^bThe MELD scale is continuous, with 34 levels ranging between 6 and 40 (scores above 40 are categorized as 40). Donor organs usually do not become available unless the MELD score exceeds 20. ^cPatients with stage T2 hepatocellular carcinoma receive 22 disease-specific points. ^dCreatinine is included because renal function is a validated predictor of survival in patients with liver disease. For adults undergoing dialysis twice a week, the creatinine in the equation is set to 4 mg/100 mL. ^eFor children <18 years of age, the Pediatric End-Stage Liver Disease (PELD) scale is used. This scale is based on albumin, bilirubin, INR, growth failure, and age. Status 1 is retained.

decisive determinant, and to eliminate both the subjectivity inherent in the CTP scoring system (presence and degree of ascites and hepatic encephalopathy) and the differences in waiting times among different regions of the country. Data indicate that liver recipients with MELD scores <15 experienced higher posttransplantation mortality rates than similarly classified patients who remained on the waiting list. This observation led to the modification of UNOS policy to allocate donor organs to candidates with MELD scores exceeding 15 within the local or regional procurement organization before offering the organ to local patients whose scores are <15. In 2016, the MELD score was modified to incorporate serum sodium, another important predictor of survival in liver transplantation candidates (the MELD-Na score).

The highest priority (status 1) continues to be reserved for patients with fulminant hepatic failure or primary graft nonfunction. Because candidates for liver transplantation who have HCC may not be sufficiently compensated to compete for donor organs based on urgency criteria alone and because protracted waiting for deceased-donor organs often results in tumor growth beyond acceptable limits for transplantation, such patients are assigned disease-specific MELD points (Table 345-3). Other disease-specific MELD exceptions include portopulmonary hypertension, hepatopulmonary syndrome, familial amyloid polyneuropathy, primary hyperoxaluria (necessitating liver-kidney transplantation), cystic fibrosis liver disease, and highly selected cases of hilar cholangiocarcinoma.

LIVING DONOR TRANSPLANTATION

Occasionally, especially for liver transplantation in children, one deceased-donor organ can be split between two recipients (one adult and one child). A more viable alternative, transplantation of the right lobe of the liver from a healthy adult donor into an adult recipient, has gained increased popularity. Living-donor transplantation of the left lobe (left lateral segment), introduced in the early 1990s to alleviate the extreme shortage of donor organs for small children, accounts currently for approximately one-third of all liver transplantation procedures in children. Driven by the shortage of deceased-donor organs, living-donor transplantation involving the more sizable right lobe is being considered with increasing frequency in adults; however, living-donor liver transplantation cannot be expected to solve the donor organ shortage; 524 such procedures were done in 2019, representing only ~4% of all liver transplant operations done in the United States.

Living-donor transplantation can reduce waiting time and cold ischemia time; is done under elective, rather than emergency, circumstances; and may be lifesaving in recipients who cannot afford to wait for a deceased donor. The downside, of course, is the risk to the healthy donor (a mean of 10 weeks of medical disability; biliary complications in ~5%; postoperative complications such as wound infection,

small-bowel obstruction, and incisional hernias in 9–19%; and even, in 0.2–0.4%, death) as well as the increased frequency of biliary (15–32%) and vascular (10%) complications in the recipient. Potential donors must participate voluntarily without coercion, and transplantation teams should go to great lengths to exclude subtle coercive or inappropriate psychological factors as well as outline carefully to both donor and recipient the potential benefits and risks of the procedure. Donors for the procedure should be 18–60 years old; have a compatible blood type with the recipient; have no chronic medical problems or history of major abdominal surgery; be related genetically or emotionally to the recipient; and pass an exhaustive series of clinical, biochemical, and serologic evaluations to unearth disqualifying medical disorders. The recipient should meet the same UNOS criteria for liver transplantation as recipients of a deceased donor allograft.

SURGICAL TECHNIQUE

Removal of the recipient's native liver is technically difficult, particularly in the presence of portal hypertension with its associated collateral circulation and extensive varices and especially in the presence of scarring from previous abdominal operations. The combination of portal hypertension and coagulopathy (elevated prothrombin time and thrombocytopenia) may translate into large blood product transfusion requirements. After the portal vein and infrahepatic and suprahepatic inferior vena cava are dissected, the hepatic artery and common bile duct are dissected. Then the native liver is removed and the donor organ inserted. During the anhepatic phase, coagulopathy, hypoglycemia, hypocalcemia, and hypothermia are encountered and must be managed by the anesthesiology team. Caval, portal vein, hepatic artery, and bile duct anastomoses are performed in succession, the last by end-to-end suturing of the donor and recipient common bile ducts (Fig. 345-1) or by choledochojejunostomy to a Roux-en-Y loop if the recipient common bile duct cannot be used for reconstruction (e.g., in sclerosing cholangitis). A typical transplant operation lasts 8 h, with a range of 6–18 h. Because of excessive bleeding, large volumes of blood, blood products, and volume expanders may be required during surgery; however, blood requirements have fallen sharply with improvements in surgical technique, blood-salvage interventions, and experience.

As noted above, emerging alternatives to orthotopic liver transplantation include split-liver grafts, in which one donor organ is divided and inserted into two recipients, and living-donor procedures, in which part of the left (for children), the left (for children or small adults), or

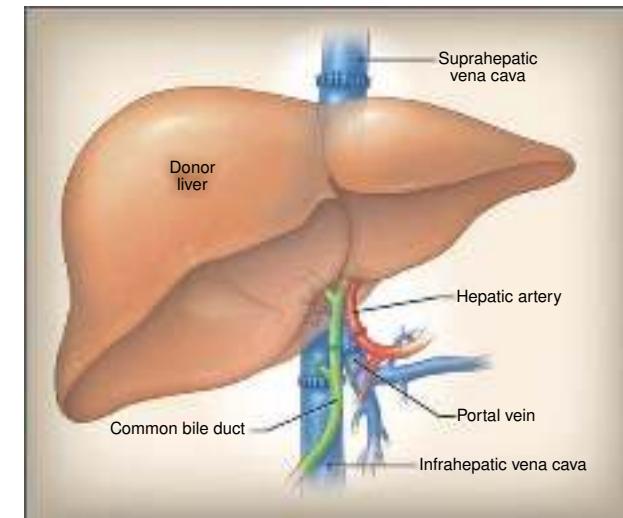


FIGURE 345-1 The anastomoses in orthotopic liver transplantation. The anastomoses are performed in the following sequence: (1) suprahepatic and infrahepatic vena cava, (2) portal vein, (3) hepatic artery, and (4) common bile duct-to-duct anastomosis. (From JL Dienstag, AB Cosimi: Liver transplantation—a vision realized. *N Engl J Med* 367:1483, 2012. Copyright © 2012 Massachusetts Medical Society. Reprinted with permission from Massachusetts Medical Society.)

the right (for adults) lobe of the liver is harvested from a living donor for transplantation into the recipient. In the adult procedure, once the right lobe is removed from the donor, the donor right hepatic vein is anastomosed to the recipient right hepatic vein remnant, followed by donor-to-recipient anastomoses of the portal vein and then the hepatic artery. Finally, the biliary anastomosis is performed, duct-to-duct if practical or via Roux-en-Y anastomosis. Heterotopic liver transplantation, in which the donor liver is inserted without removal of the native liver, has met with very limited success and acceptance, except in a very small number of centers. In attempts to support desperately ill patients until a suitable donor organ can be identified, several transplantation centers are studying extracorporeal perfusion with bioartificial liver cartridges constructed from hepatocytes bound to hollow fiber systems and used as temporary hepatic-assist devices, but their efficacy remains to be established. Areas of research with the potential to overcome the shortage of donor organs include hepatocyte transplantation and xenotransplantation with genetically modified organs of nonhuman origin (e.g., swine).

POSTOPERATIVE COURSE AND MANAGEMENT

■ IMMUNOSUPPRESSIVE THERAPY

The introduction in 1980 of cyclosporine as an immunosuppressive agent contributed substantially to the improvement in survival after liver transplantation. Cyclosporine, a calcineurin inhibitor, blocks early activation of T cells and is specific for T-cell functions that result from the interaction of the T cell with its receptor and that involve the calcium-dependent signal transduction pathway. As a result, the activity of cyclosporine leads to inhibition of lymphokine gene activation, blocking interleukins 2, 3, and 4, tumor necrosis factor α , and other lymphokines. Cyclosporine also inhibits B-cell functions. This process occurs without affecting rapidly dividing cells in the bone marrow, which may account for the reduced frequency of posttransplantation systemic infections. The most common and important side effect of cyclosporine therapy is nephrotoxicity. Cyclosporine causes dose-dependent renal tubular injury and direct renal artery vasospasm. Following renal function is therefore important in monitoring cyclosporine therapy and is perhaps even a more reliable indicator than blood levels of the drug. Nephrotoxicity is reversible and can be managed by dose reduction. Other adverse effects of cyclosporine therapy include hypertension, hyperkalemia, tremor, hirsutism, glucose intolerance, and gingival hyperplasia.

Tacrolimus, a macrolide lactone antibiotic isolated from a Japanese soil fungus, *Streptomyces tsukubaensis*, has the same mechanism of action as cyclosporine but is 10–100 times more potent. Initially applied as “rescue” therapy for patients in whom rejection occurred despite the use of cyclosporine, tacrolimus was shown to be associated with a reduced frequency of acute, refractory, and chronic rejection. Although patient and graft survival are the same with these two drugs, the advantage of tacrolimus in minimizing episodes of rejection, reducing the need for additional glucocorticoid doses, and reducing the likelihood of bacterial and cytomegalovirus (CMV) infection has simplified the management of patients undergoing liver transplantation. In addition, the oral absorption of tacrolimus is more predictable than that of cyclosporine, especially during the early postoperative period when T-tube drainage interferes with the enterohepatic circulation of cyclosporine. As a result, in most transplantation centers, tacrolimus has now supplanted cyclosporine for primary immunosuppression, and many centers rely on oral rather than IV administration from the outset. For transplantation centers that prefer cyclosporine, a better-absorbed microemulsion preparation is available.

Although more potent than cyclosporine, tacrolimus is also more toxic and more likely to be discontinued for adverse events. The toxicity of tacrolimus is similar to that of cyclosporine; nephrotoxicity and neurotoxicity are the most commonly encountered adverse effects, and neurotoxicity (tremor, seizures, hallucinations, psychoses, coma) is more likely and more severe in tacrolimus-treated patients. Both drugs can cause diabetes mellitus, but tacrolimus does not cause

hirsutism or gingival hyperplasia. Because of overlapping toxicity between cyclosporine and tacrolimus, especially nephrotoxicity, and because tacrolimus reduces cyclosporine clearance, these two drugs should not be used together. Because 99% of tacrolimus is metabolized by the liver, hepatic dysfunction reduces its clearance; in primary graft nonfunction (when, for technical reasons or because of ischemic damage prior to its insertion, the allograft is defective and does not function normally from the outset), tacrolimus doses have to be reduced substantially, especially in children. Both cyclosporine and tacrolimus are metabolized by the cytochrome P450 IIIA system, and therefore, drugs that induce cytochrome P450 (e.g., phenytoin, phenobarbital, carbamazepine, rifampin) reduce available levels of cyclosporine and tacrolimus, and drugs that inhibit cytochrome P450 (e.g., erythromycin, fluconazole, ketoconazole, clotrimazole, itraconazole, verapamil, diltiazem, danazol, metoclopramide, the HIV protease inhibitor ritonavir, and the HCV protease inhibitors glecaprevir [cyclosporine only] and grazoprevir) increase cyclosporine and tacrolimus blood levels. Indeed, itraconazole is used occasionally to help boost tacrolimus levels. Like azathioprine, cyclosporine and tacrolimus appear to be associated with a risk of lymphoproliferative malignancies (see below), which may occur earlier after cyclosporine or tacrolimus than after azathioprine therapy. Because of these side effects, combinations of cyclosporine or tacrolimus with prednisone and an antimetabolite (azathioprine or mycophenolic acid, see below)—all at reduced doses—are preferable regimens for immunosuppressive therapy.

Mycophenolic acid, a nonnucleoside purine metabolism inhibitor derived as a fermentation product from several *Penicillium* species, is another immunosuppressive drug being used for patients undergoing liver transplantation. Mycophenolate has been shown to be better than azathioprine, when used with other standard immunosuppressive drugs, in preventing rejection after renal transplantation and has been adopted widely as well for use in liver transplantation. The most common adverse effects of mycophenolate are bone marrow suppression and gastrointestinal complaints.

In patients with pretransplantation renal dysfunction or renal deterioration that occurs intraoperatively or immediately postoperatively, tacrolimus or cyclosporine therapy may not be practical; under these circumstances, induction or maintenance of immunosuppression with antithymocyte globulin (ATG; thymoglobulin) or monoclonal antibodies to T cells, OKT3, may be appropriate. Therapy with these agents has been especially effective in reversing acute rejection in the post-transplantation period and is the standard treatment for acute rejection that fails to respond to methylprednisolone boluses. Available data support the use of thymoglobulin induction to delay calcineurin inhibitor use and its attendant nephrotoxicity. IV infusions of thymoglobulin may be complicated by fever and chills, which can be ameliorated by premedication with antipyretics and a low dose of glucocorticoids. Infusions of OKT3 may be complicated by fever, chills, and diarrhea or by pulmonary edema, which can be fatal. Because OKT3 is such a potent immunosuppressive agent, its use is also more likely to be complicated by opportunistic infection or lymphoproliferative disorders; therefore, because of the availability of alternative immunosuppressive drugs, OKT3 is now used sparingly.

Sirolimus, an inhibitor of the mammalian target of rapamycin (mTOR), blocks later events in T-cell activation and is approved for use in kidney transplantation, but it is not formally approved for use in liver transplant recipients because of the reported association with an increased frequency of hepatic artery thrombosis in the first month after transplantation. In patients with calcineurin inhibitor-related nephrotoxicity, conversion to sirolimus has been demonstrated to be effective in preventing rejection with accompanying improvements in renal function. Because of its profound antiproliferative effects, sirolimus has also been suggested to be a useful immunosuppressive agent in patients with a prior or current history of malignancy, such as HCC. Side effects include hyperlipidemia, peripheral edema, oral ulcers, and interstitial pneumonitis. Everolimus is a hydroxyethyl derivative of sirolimus that, when used in conjunction with low-dose tacrolimus, also provides successful protection against acute rejection, with decreased renal impairment compared to that associated with standard

2638 tacrolimus dosing. Everolimus and sirolimus share a similar adverse events profile. The most important principle of immunosuppression is that the ideal approach strikes a balance between immunosuppression and immunologic competence. In general, given sufficient immunosuppression, acute liver allograft rejection is nearly always reversible. On one hand, incompletely treated acute rejection predisposes to the development of chronic rejection, which can threaten graft survival. On the other hand, if the cumulative dose of immunosuppressive therapy is too large, the patient may succumb to opportunistic infection. In hepatitis C, pulse glucocorticoids or OKT3 use accelerates recurrent allograft hepatitis, although the routine use of DAA therapy to clear the allograft of HCV should remove this concern. Further complicating matters, acute rejection can be difficult to distinguish histologically from recurrent hepatitis C. Therefore, immunosuppressive drugs must be used judiciously, with strict attention to the infectious consequences of such therapy and careful confirmation of the diagnosis of acute rejection. In this vein, efforts have been made to minimize the use of glucocorticoids, a mainstay of immunosuppressive regimens, and steroid-free immunosuppression can be achieved in some instances. Patients who undergo liver transplantation for autoimmune diseases such as primary biliary cholangitis, autoimmune hepatitis, and primary sclerosing cholangitis are less likely to achieve freedom from glucocorticoids.

■ POSTOPERATIVE COMPLICATIONS

Complications of liver transplantation can be divided into nonhepatic and hepatic categories (Tables 345-4 and 345-5). In addition, both immediate postoperative and late complications are encountered. As a rule, patients who undergo liver transplantation have been chronically ill for protracted periods and may be malnourished and wasted. The impact of such chronic illness and the multisystem failure that accompanies liver failure continue to require attention in the postoperative period. Because of the massive fluid losses and fluid shifts that occur

TABLE 345-5 Hepatic Complications of Liver Transplantation

Hepatic Dysfunction Common after Major Surgery

Prehepatic	Pigment load Hemolysis Blood collections (hematomas, abdominal collections)
Intrahepatic	Hepatotoxic drugs and anesthesia
	Hypoperfusion (hypotension, shock, sepsis) Benign postoperative cholestasis
Posthepatic	Transfusion-associated hepatitis Exacerbation of primary hepatic disease
	Biliary obstruction ↓ Renal clearance of conjugated bilirubin (renal dysfunction)

Hepatic Dysfunction Unique to Liver Transplantation

Primary graft nonfunction	
Vascular compromise	Portal vein obstruction Hepatic artery thrombosis Anastomotic leak with intraabdominal bleeding
Bile duct disorder	Stenosis, obstruction, leak
Rejection	
Recurrent primary hepatic disease	

during the operation, patients may remain fluid overloaded during the immediate postoperative period, straining cardiovascular reserve; this effect can be amplified in the face of transient renal dysfunction and pulmonary capillary vascular permeability. Continuous monitoring of cardiovascular and pulmonary function, measures to maintain the integrity of the intravascular compartment and to treat extravascular volume overload, and scrupulous attention to potential sources and sites of infection are of paramount importance. Cardiovascular instability may also result from the electrolyte imbalance that may accompany reperfusion of the donor liver as well as from restoration of systemic vascular resistance following implantation. Pulmonary function may be compromised further by paralysis of the right hemidiaphragm associated with phrenic nerve injury. The hyperdynamic state with increased cardiac output that is characteristic of patients with liver failure reverses rapidly after successful liver transplantation.

Other immediate management issues include renal dysfunction. Prerenal azotemia, acute kidney injury associated with hypoperfusion (acute tubular necrosis), and renal toxicity caused by antibiotics, tacrolimus, or cyclosporine are encountered frequently in the postoperative period, sometimes necessitating dialysis. Hemolytic-uremic syndrome can be associated with cyclosporine, tacrolimus, or OKT3. Occasionally, postoperative intraperitoneal bleeding may be sufficient to increase intraabdominal pressure, which, in turn, may reduce renal blood flow; this effect is rapidly reversible when abdominal distention is relieved by exploratory laparotomy to identify and ligate the bleeding site and to remove intraperitoneal clot.

Anemia may also result from acute upper gastrointestinal bleeding or from transient hemolytic anemia, which may be autoimmune, especially when blood group O livers are transplanted into blood group A or B recipients. This autoimmune hemolytic anemia is mediated by donor intrahepatic lymphocytes that recognize red blood cell A or B antigens on recipient erythrocytes. Transient in nature, this process resolves once the donor liver is repopulated by recipient bone marrow-derived lymphocytes; the hemolysis can be treated by transfusing blood group O red blood cells and/or by administering higher doses of glucocorticoids. Transient thrombocytopenia is also commonly encountered. Aplastic anemia, a late occurrence, is rare but has been reported in almost 30% of patients who underwent liver transplantation for acute, severe hepatitis of unknown cause.

TABLE 345-4 Nonhepatic Complications of Liver Transplantation

CATEGORY	COMPLICATION
Cardiovascular instability	Arrhythmias Congestive heart failure Cardiomyopathy
Pulmonary compromise	Pneumonia Pulmonary capillary vascular permeability Fluid overload
Renal dysfunction	Prerenal azotemia Hypoperfusion injury (acute tubular necrosis) Drug nephrotoxicity ↓ Renal blood flow secondary to ↑ intraabdominal pressure
Hematologic	Anemia secondary to gastrointestinal and/or intraabdominal bleeding Hemolytic anemia, aplastic anemia Thrombocytopenia
Infection	Bacterial: early, common postoperative infections Fungal/parasitic: late, opportunistic infections Viral: late, opportunistic infections, recurrent hepatitis
Neuropsychiatric	Seizures Metabolic encephalopathy Depression Difficult psychosocial adjustment
Diseases of donor	Infectious Malignant
Malignancy	B-cell lymphoma (posttransplantation lymphoproliferative disorders) De novo neoplasms (particularly squamous cell skin carcinoma)

Bacterial, fungal, or viral infections are common and may be life-threatening postoperatively. Early after transplant surgery, common postoperative infections predominate—pneumonia, wound infections, infected intraabdominal collections, urinary tract infections, and IV line infections—rather than opportunistic infections; these infections may involve the biliary tree and liver as well. Beyond the first postoperative month, the toll of immunosuppression becomes evident, and opportunistic infections—CMV, herpes viruses, fungal infections (*Aspergillus*, *Candida*, cryptococcal disease), mycobacterial infections, parasitic infections (*Pneumocystis*, *Toxoplasma*), and bacterial infections (*Nocardia*, *Legionella*, *Listeria*)—predominate. Rarely, early infections represent those transmitted with the donor liver, either infections present in the donor or infections acquired during procurement processing. De novo viral hepatitis infections acquired from the donor organ or, almost unheard of now, from transfused blood products occur after typical incubation periods for these agents (well beyond the first month). Obviously, infections in an immunosuppressed host demand early recognition and prompt management; prophylactic antibiotic therapy is administered routinely in the immediate postoperative period. Use of sulfamethoxazole with trimethoprim reduces the incidence of postoperative *Pneumocystis jirovecii* pneumonia. Antiviral prophylaxis for CMV with ganciclovir should be administered in patients at high risk (e.g., when a CMV-seropositive donor organ is implanted into a CMV-seronegative recipient).

Neuropsychiatric complications include seizures (commonly associated with cyclosporine and tacrolimus toxicity), metabolic encephalopathy, depression, and difficult psychosocial adjustment. Rarely, diseases are transmitted by the allograft from the donor to the recipient. In addition to viral and bacterial infections, malignancies of donor origin have occurred. Posttransplantation lymphoproliferative disorders, especially B-cell lymphoma, are a recognized complication associated with immunosuppressive drugs such as azathioprine, tacrolimus, and cyclosporine (see above). Epstein-Barr virus has been shown to play a contributory role in some of these tumors, which may regress when immunosuppressive therapy is reduced. De novo neoplasms appear at increased frequency after liver transplantation, particularly squamous cell carcinomas of the skin. Routine screening should be performed.

Long-term complications after liver transplantation attributable primarily to immunosuppressive medications include diabetes mellitus and osteoporosis (associated with glucocorticoids and calcineurin inhibitors) as well as hypertension, hyperlipidemia, and chronic renal insufficiency (associated with cyclosporine and tacrolimus). Monitoring and treating these disorders are routine components of posttransplantation care; in some cases, they respond to changes in immunosuppressive regimen, while in others, specific treatment of the disorder is introduced. Data from a large U.S. database showed that the prevalence of renal failure was 18% at year 5 and 25% at year 10 after liver transplantation. Similarly, the high frequency of diabetes, hypertension, hyperlipidemia, obesity, and the metabolic syndrome renders patients susceptible to cardiovascular disease after liver transplantation; although hepatic complications account for most of the mortality after liver transplantation, renal failure and cardiovascular disease are the other leading causes of late mortality after liver transplantation.

■ HEPATIC COMPLICATIONS

Hepatic dysfunction after liver transplantation is similar to the hepatic complications encountered after major abdominal and cardiothoracic surgery; however, in addition, hepatic complications include primary graft failure, vascular compromise, failure or stricture of the biliary anastomoses, and rejection. As in nontransplantation surgery, postoperative jaundice may result from prehepatic, intrahepatic, and posthepatic sources. *Prehepatic* sources represent the massive hemoglobin pigment load from transfusions, hemolysis, hematomas, ecchymoses, and other collections of blood. *Early intrahepatic* liver injury includes effects of hepatotoxic drugs and anesthesia; hypoperfusion injury associated with hypotension, sepsis, and shock; and benign postoperative cholestasis. *Late intrahepatic* sources of liver injury include exacerbation of primary disease. *Posthepatic* sources of hepatic dysfunction include biliary obstruction and reduced renal clearance of conjugated bilirubin.

Hepatic complications unique to liver transplantation include primary graft failure associated with ischemic injury to the organ during harvesting; vascular compromise associated with thrombosis or stenosis of the portal vein or hepatic artery anastomoses; vascular anastomotic leak; stenosis, obstruction, or leakage of the anastomosed common bile duct; recurrence of primary hepatic disorder (see below); and rejection.

■ TRANSPLANT REJECTION

Despite the use of immunosuppressive drugs, rejection of the transplanted liver still occurs in a proportion of patients, beginning 1–2 weeks after surgery. Clinical signs suggesting rejection are fever, right upper quadrant pain, and reduced bile pigment and volume. Leukocytosis may occur, but the most reliable indicators are increases in serum bilirubin and aminotransferase levels. Because these tests lack specificity, distinguishing among rejection, biliary obstruction, primary graft nonfunction, vascular compromise, viral hepatitis, CMV infection, drug hepatotoxicity, and recurrent primary disease may be difficult. Radiographic visualization of the biliary tree and/or percutaneous liver biopsy often help to establish the correct diagnosis. Morphologic features of acute rejection include a mixed portal cellular infiltrate, bile duct injury, and/or endothelial inflammation ("endothelitis"); some of these findings are reminiscent of graft-versus-host disease, primary biliary cholangitis, or recurrent allograft hepatitis C. As soon as transplant rejection is suspected, treatment consists of IV methylprednisolone in repeated boluses; if this fails to abort rejection, many centers use thymoglobulin or OKT3. Caution should be exercised when managing acute rejection with pulse glucocorticoids or OKT3 in patients with HCV infection, because of the high risk of triggering recurrent allograft hepatitis C. The availability of DAAs for HCV has effectively obviated this concern.

Chronic rejection is a relatively rare outcome that can follow repeated bouts of acute rejection or that occurs unrelated to preceding rejection episodes. Morphologically, chronic rejection is characterized by progressive cholestasis, focal parenchymal necrosis, mononuclear infiltration, vascular lesions (intimal fibrosis, subintimal foam cells, fibrinoid necrosis), and fibrosis. This process may be reflected as ductopenia—the vanishing bile duct syndrome, which is more common in patients undergoing liver transplantation for autoimmune liver disease. Reversibility of chronic rejection is limited; in patients with therapy-resistant chronic rejection, retransplantation has yielded encouraging results.

OUTCOME

■ SURVIVAL

The survival rate for patients undergoing liver transplantation has improved steadily since 1983. One-year survival rates have increased from ~70% in the early 1980s to 85–90% from 2003 to the present time. Currently, the 5-year survival rate exceeds 60%. An important observation is the relationship between clinical status before transplantation and outcome. For patients who undergo liver transplantation when their level of compensation is high (e.g., still working or only partially disabled), a 1-year survival rate of >85% is common. For those whose level of decompensation mandates continuous in-hospital care prior to transplantation, the 1-year survival rate is ~70%, whereas for those who are so decompensated that they require life support in an intensive care unit, the 1-year survival rate is ~50%. Since the adoption by UNOS in 2002 of the MELD system for organ allocation, posttransplantation survival has been found to be affected adversely for candidates with MELD scores >25, considered high disease severity. Thus, irrespective of allocation scheme, high disease severity before transplantation corresponds to diminished posttransplantation survival. Another important distinction in survival has been drawn between high- and low-risk patient categories. For patients who do not fit any "high-risk" designations, 1- and 5-year survival rates of 85 and 80%, respectively, have been recorded. In contrast, among patients in high-risk categories—cancer, fulminant hepatitis, age >65, concurrent renal failure, respiratory dependence, portal vein thrombosis, and history of a portacaval shunt or multiple right upper quadrant operations—survival statistics fall into the range of 60% at 1 year and 35% at 5 years. Survival after

retransplantation for primary graft nonfunction is ~50%. Causes of failure of liver transplantation vary with time. Failures within the first 3 months result primarily from technical complications, postoperative infections, and hemorrhage. Transplant failures after the first 3 months are more likely to result from infection, rejection, or recurrent disease (such as malignancy or viral hepatitis).

■ RECURRENCE OF PRIMARY DISEASE

Features of autoimmune hepatitis, primary sclerosing cholangitis, and primary biliary cholangitis overlap with those of rejection or post-transplantation bile duct injury. Whether autoimmune hepatitis and sclerosing cholangitis recur after liver transplantation is controversial; data supporting recurrent autoimmune hepatitis (in up to one-third of patients in some series) are more convincing than those supporting recurrent sclerosing cholangitis. Similarly, reports of recurrent primary biliary cholangitis after liver transplantation have appeared; however, the histologic features of primary biliary cholangitis and chronic rejection are virtually indistinguishable and occur as frequently in patients with primary biliary cholangitis as in patients undergoing transplantation for other reasons. The presence of a florid inflammatory bile duct lesion is highly suggestive of the recurrence of primary biliary cholangitis, but even this lesion can be observed in acute rejection. Hereditary disorders such as Wilson's disease and α_1 antitrypsin deficiency have not recurred after liver transplantation; however, recurrence of disordered iron metabolism has been observed in some patients with hemochromatosis. Hepatic vein thrombosis (Budd-Chiari syndrome) may recur; this can be minimized by treating underlying myeloproliferative disorders and by anticoagulation. Because cholangiocarcinoma recurs almost invariably, few centers now offer transplantation to such patients; however, a few highly selected patients with operatively confirmed stage I or II cholangiocarcinoma who undergo liver transplantation combined with neoadjuvant chemoradiation may experience excellent outcomes. In patients with intrahepatic HCC who meet criteria for transplantation, 1- and 5-year survivals are similar to those observed in patients undergoing liver transplantation for nonmalignant disease. Finally, metabolic disorders such as NAFLD recur frequently, especially if the underlying metabolic predisposition is not altered. The metabolic syndrome occurs commonly after liver transplantation as a result of recurrent NAFLD, immunosuppressive medications, and/or, in patients with hepatitis C related to the impact of HCV infection on insulin resistance, diabetes and fatty liver.

Hepatitis A can recur after transplantation for fulminant hepatitis A, but such acute reinfection has no serious clinical sequelae. In fulminant hepatitis B, recurrence is not the rule; however, in the absence of any prophylactic measures, hepatitis B usually recurs after transplantation for end-stage chronic hepatitis B. Before the introduction of prophylactic antiviral therapy, immunosuppressive therapy sufficient to prevent allograft rejection led inevitably to marked increases in hepatitis B viremia, regardless of pretransplantation levels. Overall graft and patient survival were poor, and some patients experienced a rapid recapitulation of severe injury—severe chronic hepatitis or even fulminant hepatitis—after transplantation. Also recognized in the era before availability of antiviral regimens was *fibrosing cholestatic hepatitis*, rapidly progressive liver injury associated with marked hyperbilirubinemia, substantial prolongation of the prothrombin time (both out of proportion to relatively modest elevations of aminotransferase activity), and rapidly progressive liver failure. This lesion has been suggested to represent a “choking off” of the hepatocyte by an overwhelming density of HBV proteins. Complications such as sepsis and pancreatitis were also observed more frequently in patients undergoing liver transplantation for hepatitis B prior to the introduction of antiviral therapy. The introduction of long-term prophylaxis with HBIG revolutionized liver transplantation for chronic hepatitis B. Preoperative hepatitis B vaccination, preoperative or postoperative interferon (IFN) therapy, or short-term (≤ 2 months) HBIG prophylaxis has not been shown to be effective, but a retrospective analysis of data from several hundred European patients followed for 3 years after transplantation has shown that long-term (≥ 2 months) prophylaxis with HBIG is associated with a

lowering of the risk of HBV reinfection from ~75 to 35% and a reduction in mortality from ~50 to 20%.

As a result of long-term HBIG use following liver transplantation for chronic hepatitis B, similar improvements in outcome have been observed in the United States, with 1-year survival rates between 75 and 90%. Currently, with HBIG prophylaxis, the outcome of liver transplantation for chronic hepatitis B is indistinguishable from that for chronic liver disease unassociated with chronic hepatitis B; essentially, medical concerns regarding liver transplantation for chronic hepatitis B have been eliminated. Passive immunoprophylaxis with HBIG is begun during the anhepatic stage of surgery, repeated daily for the first 6 postoperative days, and then continued with infusions that are given either at regular intervals of 4–6 weeks or, alternatively, when anti-hepatitis B surface (HBs) levels fall below a threshold of 100 mIU/mL. The current approach in most centers is to continue HBIG indefinitely, which can add ~\$20,000 per year to the cost of care; some centers are evaluating regimens that shift to less frequent administration or to IM administration in the late posttransplantation period or, in low-risk patients, maintenance with antiviral therapy (see below) alone. Still, “breakthrough” HBV infection occasionally occurs.

Further improving the outcome of liver transplantation for chronic hepatitis B is the current availability of such antiviral drugs as entecavir, tenofovir disoproxil fumarate, and tenofovir alafenamide (Chap. 341). When these drugs are administered to patients with decompensated liver disease, a proportion improves sufficiently to postpone imminent liver transplantation. In addition, antiviral therapy can be used to prevent recurrence of HBV infection when administered *prior to* transplantation; to treat hepatitis B that recurs *after* transplantation, including in patients who break through HBIG prophylaxis; and to reverse the course of otherwise fatal fibrosing cholestatic hepatitis. Clinical trials have shown that entecavir or tenofovir antiviral therapy reduces the level of HBV replication substantially, sometimes even resulting in clearance of hepatitis B surface antigen (HBsAg); reduces alanine aminotransferase (ALT) levels; and improves histologic features of necrosis and inflammation. Currently, most liver transplantation centers combine HBIG plus one of the high-barrier-to-resistance oral nucleoside (entecavir) or nucleotide analogues (tenofovir). In low-risk patients with no detectable hepatitis B viremia at the time of transplantation, a number of clinical trials have suggested that antiviral prophylaxis can suffice, without HBIG or with a finite duration of HBIG, to prevent recurrent HBV infection of the allograft. In patients documented at the time of liver transplantation to have undetectable HBV DNA in serum and covalently closed circular DNA in the liver (i.e., with low risk for recurrence of HBV infection), a preliminary clinical trial suggested that, after receipt of 5 years of combined therapy, both HBIG and oral-agent therapy can be withdrawn sequentially (over two 6-month periods) with a success rate, as monitored over a median of 6 years after withdrawal, of 90% and an anti-HBs seroconversion rate of 60% (despite transient reappearance of HBV DNA and/or HBsAg in some of these patients).

Antiviral prophylactic approaches applied to patients undergoing liver transplantation for chronic hepatitis B are being used as well for patients without hepatitis B who receive organs from donors with antibody to hepatitis B core antigen (anti-HBc) but do not have HBsAg. Patients who undergo liver transplantation for chronic hepatitis B plus D are less likely to experience recurrent liver injury than patients undergoing liver transplantation for hepatitis B alone; still, such co-infected patients would also be offered standard posttransplantation prophylactic therapy for hepatitis B.

Until recently, the most common indication for liver transplantation was end-stage liver disease resulting from chronic hepatitis C. For patients undergoing liver transplantation for hepatitis C, because of an aggressive natural history of recurrent allograft hepatitis C, graft and patient survival were diminished substantially compared to other indications for transplantation.

The approval over the last decade of several DAA agents and of IFN-free DAA regimens against HCV has had a major impact on the management and outcome of both pretransplantation and posttransplantation HCV infection. Such therapeutic approaches (1) permit

the clearance of viremia in a substantial proportion of decompensated cirrhotics, thereby preventing recurrent allograft infection and even improving the clinical status of most of these patients, delaying or obviating the need for liver replacement; and (2) achieve sustained virologic responses in a much higher proportion of persons with allograft HCV infection, because of improvements in antiviral treatment efficacy and tolerability. Ideally, patients should be treated prior to liver transplantation. This approach has already reduced the numbers of patients referred for liver transplantation and led to delisting of others. A concern, however, is that eradication of HCV infection will reduce the MELD score and lower the priority for a donor organ in some patients who still require transplantation because of continued hepatic decompensation and profound reduction in quality of life. In addition, elimination of HCV infection prior to transplantation would disqualify such patients from accepting donor livers from persons with HCV infection, contracting the potential donor pool and limiting accessibility to donor organs and timely transplantation. Therefore, consideration should be given to postponing DAA therapy in patients with high-MELD HCV-associated end-stage liver disease until after liver transplantation; however, a distinct threshold at which to treat pretransplantation or posttransplantation has not yet been established. Regardless, the approach to treatment should be individualized thoughtfully for each patient, based on such factors as MELD score, time anticipated prior to availability of a donor organ, relative clinical stability, and comorbidities.

DAA combinations that have been used successfully against allograft HCV include ledipasvir, sofosbuvir, and ribavirin; velpatasvir, sofosbuvir, and ribavirin; and grazoprevir and pibrentasvir. (For updated guidelines, see www.hcvguidelines.org.) In patients with recurrent HCV infection after liver transplantation, each of these regimens has yielded response rates approaching those seen in compensated nontransplant patient populations.

A small number of allograft recipients have historically succumbed to early HCV-associated liver injury, and a syndrome reminiscent of fibrosing cholestatic hepatitis (see above) has been observed rarely. Currently, however, the routine use of DAA regimens early after transplantation, before the onset of these variant presentations, has already had a profound impact on the frequency of severe recurrent allograft hepatitis C.

Patients who undergo liver transplantation for end-stage alcohol-associated cirrhosis are at risk of resorting to drinking again after transplantation, a potential source of recurrent alcohol-associated liver injury. Currently, alcohol-associated liver disease is the most common indication for liver transplantation, accounting for 30% of all liver transplantation procedures, and most transplantation centers screen candidates carefully for predictors of continued abstinence. Recidivism is more likely in patients whose sobriety prior to transplantation was <6 months. For abstinent patients with alcohol-associated cirrhosis, liver transplantation can be undertaken successfully, with outcomes comparable to those for other categories of patients with chronic liver disease, when coordinated by a team approach that includes substance abuse counseling.

■ POSTTRANSPLANTATION QUALITY OF LIFE

Full rehabilitation is achieved in most patients who survive the early postoperative months and escape chronic rejection or unmanageable infection. Psychosocial maladjustment interferes with medical compliance in a small number of patients, but most manage to adhere to immunosuppressive regimens, which must be continued indefinitely. In one study, 85% of patients who survived their transplant operations returned to gainful activities. In fact, some women have conceived and carried pregnancies to term after transplantation without demonstrable injury to their infants.

■ FURTHER READING

AASLD/IDSA HCV G P : Hepatitis C guidance 2019 update: American Association for the Study of Liver Diseases-Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. *Hepatology* 71:686,

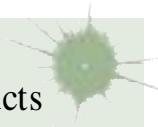
2020. (Updated regularly, available at <http://www.hcvguidelines.org>) Accessed August 24, 2021.

- C TG et al: Improved graft survival after liver transplantation for recipients with hepatitis C virus in the direct-acting antiviral era. *Liver Transpl* 25:598, 2019.
- E A S L : EASL clinical practice guidelines: Liver transplantation. *J Hepatol* 64:433, 2016.
- F J et al: Outcomes including liver histology after liver transplantation for chronic hepatitis B using oral antiviral therapy alone. *Liver Transpl* 21:1504, 2015.
- G D et al: Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis or liver failure on the waitlist for liver transplantation. *Gastroenterology* 152:1090, 2017.
- K PY et al: An interferon-free antiviral regimen for HCV after liver transplantation. *N Engl J Med* 371:2375, 2014.
- L I et al: Complete hepatitis B virus prophylaxis withdrawal in hepatitis B surface antigen-positive liver transplant recipients after long-term minimal immunosuppression. *Liver Transpl* 22:1205, 2016.
- L MR et al: Long-term management of the successful adult liver transplant: 2012 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation. *Liver Transpl* 19:3, 2013.
- M M et al: Ledipasvir and sofosbuvir plus ribavirin in patients with genotype 1 or 4 hepatitis C virus infection and advanced liver disease: A multicentre, open-label, randomised, phase 2 trial. *Lancet Infect Dis* 16:685, 2016.
- M P et al: Evaluation for liver transplantation in adults: 2013 practice guideline by the American Association for the Study of Liver Diseases and the American Society of Transplantation. *Hepatology* 59:1145, 2014.
- R N et al: Glecaprevir/pibrentasvir treatment in liver or kidney transplant patients with hepatitis C virus infection. *Hepatology* 68:1298, 2018.

346

Diseases of the Gallbladder and Bile Ducts

Norton J. Greenberger*, Gustav Paumgartner,
Daniel S. Pratt



PHYSIOLOGY OF BILE PRODUCTION AND FLOW

■ BILE SECRETION AND COMPOSITION

Bile formed in hepatocytes is secreted into a complex network of canaliculi, small bile ductules, and larger bile ducts that run with lymphatics and branches of the portal vein and hepatic artery in portal tracts situated between hepatic lobules. These interlobular bile ducts coalesce to form larger septal bile ducts that join to form the right and left hepatic ducts, which in turn, unite to form the common hepatic duct. The common hepatic duct is joined by the cystic duct of the gallbladder to form the common bile duct (CBD), which enters the duodenum (often after joining the main pancreatic duct) through the ampulla of Vater.

Hepatic bile is an isotonic fluid with an electrolyte composition resembling blood plasma. The electrolyte composition of gallbladder bile differs from that of hepatic bile because most of the inorganic anions, chloride, and bicarbonate have been removed by reabsorption across the gallbladder epithelium. As a result of water reabsorption, total solute concentration of bile increases from 3–4 g/dL in hepatic bile to 10–15 g/dL in gallbladder bile.

Major solute components of bile by moles percent include bile acids (80%), phospholipids (lecithins, cephalins, and sphingomyelin) (16%), and unesterified cholesterol (4.0%). In the lithogenic state, the cholesterol value can be as high as 8–10%. Other constituents include conjugated bilirubin; proteins (all immunoglobulins, albumin, metabolites of hormones, and other proteins metabolized in the liver); electrolytes; mucus; heavy metals; and, often, drugs and their metabolites.

The total daily basal secretion of hepatic bile is ~500–600 mL. Many substances taken up or synthesized by the hepatocyte are secreted into the bile canalliculi. The canalicular membrane forms microvilli and is associated with microfilaments of actin, microtubules, and other contractile elements. Prior to their secretion into the bile, many substances are taken up into the hepatocyte, while others, such as phospholipids, a portion of primary bile acids, and some cholesterol, are synthesized *de novo* in the hepatocyte. Three mechanisms are important in regulating bile flow: (1) active transport of bile acids from hepatocytes into the bile canalliculi, (2) active transport of other organic anions, and (3) cholangiocellular secretion. The last is a secretin-mediated and cyclic AMP-dependent mechanism that results in the secretion of a bicarbonate-rich fluid into the bile ducts.

Active vectorial trans-hepatocellular movement of bile acids from the portal blood into the bile canalliculi is driven by a set of transport systems at the basolateral (sinusoidal) and the canalicular apical plasma membrane domains of the hepatocyte. Two sinusoidal bile salt uptake systems have been cloned in humans, the Na^+ /taurocholate cotransporter (NTCP, SLC10A1) and the organic anion-transporting proteins (OATP1B1/1B3), which also transport a large variety of non-bile salt organic anions. Several ATP-dependent canalicular transport systems, “export pumps” (ATP-binding cassette transport proteins, also known as ABC transporters), have been identified, the most important of which are the bile salt export pump (BSEP, ABCB1); the anionic conjugate export pump (MRP2, ABCC2), which mediates the canalicular excretion of various amphiphilic conjugates formed by phase II conjugation (e.g., bilirubin mono- and diglucuronides and drugs); the multidrug export pump (MDR1, ABCB1) for hydrophobic cationic compounds; and the phospholipid export pump (MDR3, ABCB4). Two hemitransporters, ABCG5/G8, functioning as a couple, constitute the canalicular cholesterol and phytosterol transporter. FIC1 (ATP8B1) is an aminophospholipid transferase (“flippase”) essential for maintaining the lipid asymmetry of the canalicular membrane. The canalicular membrane also contains ATP-independent transport systems such as the Cl/HCO_3^- anion exchanger isoform 2 (AE2, SLC4A2) for canalicular bicarbonate secretion. For most of these transporters, genetic defects have been identified that are associated with various forms of cholestasis or defects of biliary excretion. FIC1 (ATP8B1) is defective in progressive familial intrahepatic cholestasis type 1 (PFIC1) and benign recurrent intrahepatic cholestasis type 1 (BRIC1) and results in ablation of all other ATP-dependent transporter functions. BSEP (ABCB11) is defective in PFIC2 and BRIC2. Mutations of MRP2 (ABCC2) cause the Dubin-Johnson syndrome, an inherited form of conjugated hyperbilirubinemia (Chap. 338). A defective MDR3 (ABCB4) results in PFIC3. ABCG5/G8, the canalicular half transporters for cholesterol and other neutral sterols, are defective in sitosterolemia. The cystic fibrosis transmembrane regulator (CFTR, ABCC7), located on bile duct epithelial cells but not on canalicular membranes, is defective in cystic fibrosis, which is associated with impaired cholangiocellular pH regulation during ductular bile formation and chronic cholestatic liver disease, occasionally resulting in biliary cirrhosis.

THE BILE ACIDS

The primary bile acids, cholic acid and chenodeoxycholic acid (CDCA), are synthesized in hepatocytes from cholesterol, conjugated with glycine or taurine, and secreted into the bile canalculus. Secondary bile acids, including deoxycholate and lithocholate, are formed in the colon as bacterial metabolites of the primary bile acids. However, lithocholic acid is much less efficiently absorbed from the colon than deoxycholic acid. Another secondary bile acid, found in low concentration, is ursodeoxycholic acid (UDCA), a stereoisomer of CDCA. In healthy subjects, the ratio of glycine to taurine conjugates in bile is ~3:1.

Bile acids are detergent-like molecules that in aqueous solutions and above a critical concentration of ~2 mM form molecular aggregates called *micelles*. Cholesterol alone is sparingly soluble in aqueous environments, and its solubility in bile depends on both the total lipid concentration and the relative molar percentages of bile acids and lecithin. Normal ratios of these constituents favor the formation of solubilizing *mixed micelles*, while abnormal ratios promote the precipitation of cholesterol crystals in bile via an intermediate liquid crystal phase.

In addition to facilitating the biliary excretion of cholesterol, bile acids facilitate the normal intestinal absorption of dietary fats, mainly cholesterol, and fat-soluble vitamins, via a micellar transport mechanism (Chap. 325). Bile acids also serve as a major physiologic driving force for hepatic bile flow and aid in water and electrolyte transport in the small bowel and colon.

Bile acids also function as hormones binding to nuclear (farnesoid X receptor [FXR]) and G protein-coupled (TGR5) receptors that regulate bile acid metabolism and their enterohepatic circulation.

ENTEROHEPATIC CIRCULATION

Bile acids are efficiently conserved under normal conditions. Unconjugated, and to a lesser degree also conjugated, bile acids are absorbed by *passive diffusion* along the entire gut. Quantitatively much more important for bile salt recirculation, however, is the *active transport* mechanism for conjugated bile acids in the distal ileum (Chap. 325). The reabsorbed bile acids enter the portal bloodstream and are taken up rapidly by hepatocytes, reconjugated, and resecreted into bile (enterohepatic circulation).

The normal bile acid pool size is ~2–4 g. During digestion of a meal, the bile acid pool undergoes at least one or more enterohepatic cycles, depending on the size and composition of the meal. Normally, the bile acid pool circulates ~5–10 times daily. Intestinal reabsorption of the pool is ~95% efficient; therefore, daily fecal loss of bile acids is in the range of 0.2–0.4 g. In the steady state, this fecal loss is compensated by an equal daily synthesis of bile acids by the liver, and thus, the size of the bile acid pool is maintained. Bile acids in the intestine stimulate the release of fibroblast growth factor 19 (FGF19), which suppresses the hepatic synthesis of bile acids from cholesterol by inhibiting the rate-limiting enzyme cytochrome P450 7A1 (CYP7A1). FGF19 also promotes gallbladder relaxation. While the loss of bile salts in stool is usually matched by increased hepatic synthesis, the maximum rate of synthesis is ~5 g/d, which may be insufficient to replete the bile acid pool size when there is pronounced impairment of intestinal bile salt reabsorption.

The expression of ABC transporters in the enterohepatic circulation and of the rate-limiting enzymes of bile acid and cholesterol synthesis are regulated in a coordinated fashion by nuclear receptors, which are ligand-activated transcription factors. The hepatic BSEP (ABCB11) is upregulated by the FXR that also represses bile acid synthesis. The expression of the cholesterol transporter, ABCG5/G8, is upregulated by the liver X receptor (LXR), which is an oxysterol sensor.

GALLBLADDER AND SPHINCTERIC FUNCTIONS

In the fasting state, the sphincter of Oddi (SOD) offers a high-pressure zone of resistance to bile flow from the CBD into the duodenum. Its tonic contraction serves to (1) prevent reflux of duodenal contents into the pancreatic and bile ducts and (2) promote filling of the gallbladder. The major factor controlling the evacuation of the gallbladder is the peptide hormone cholecystokinin (CCK), which is released from the duodenal mucosa in response to the ingestion of fats and amino acids. CCK produces (1) powerful contraction of the gallbladder, (2) decreased resistance of the SOD, and (3) enhanced flow of biliary contents into the duodenum.

Hepatic bile is “concentrated” within the gallbladder by energy-dependent transmucosal absorption of water and electrolytes. Almost the entire bile acid pool may be sequestered in the gallbladder following an overnight fast for delivery into the duodenum with the first meal of the day. The normal capacity of the gallbladder is ~30 mL.

■ CONGENITAL ANOMALIES

Anomalies of the biliary tract are not uncommon and include abnormalities in number, size, and shape (e.g., agenesis of the gallbladder, duplications, rudimentary or oversized “giant” gallbladders, and diverticula). *Phrygian cap* is a clinically innocuous entity in which a partial or complete septum (or fold) separates the fundus from the body. Anomalies of position or suspension are not uncommon and include left-sided gallbladder, intrahepatic gallbladder, retrodisplacement of the gallbladder, and “floating” gallbladder. The latter condition predisposes to acute torsion, volvulus, or herniation of the gallbladder.

■ GALLSTONES

Epidemiology and Pathogenesis Gallstones are quite prevalent in most Western countries. Gallstone formation increases after age 50. In the United States, the prevalence is highest in Native Americans followed by Hispanics, non-Hispanic whites, and black Americans. The prevalence is higher in women than men across all ages.

Gallstones form because of abnormal bile composition. They are divided into two major types: cholesterol stones and pigment stones. Cholesterol stones account for >90% of all gallstones in Western industrialized countries. Cholesterol gallstones usually contain >50% cholesterol monohydrate plus an admixture of calcium salts, bile pigments, proteins, and fatty acids. Pigment stones are composed primarily of calcium bilirubinate; they contain <20% cholesterol and are classified into “black” and “brown” types, the latter forming secondary to chronic biliary infection.

CHOLESTEROL STONES AND BILIARY SLUDGE Cholesterol is essentially water-insoluble and requires aqueous dispersion into either micelles or vesicles, both of which require the presence of a second lipid to solubilize the cholesterol. Cholesterol and phospholipids are secreted into bile as unilamellar bilayered vesicles, which are converted into mixed micelles consisting of bile acids, phospholipids, and cholesterol by the action of bile acids. If there is an excess of cholesterol in relation to phospholipids and bile acids, unstable, cholesterol-rich vesicles remain, which aggregate into large multilamellar vesicles from which cholesterol crystals precipitate (Fig. 346-1).

There are several important mechanisms in the formation of lithogenic (stone-forming) bile. The most important is increased biliary secretion of cholesterol. This may occur in association with obesity, the metabolic syndrome, high-caloric and cholesterol-rich diets, or drugs (e.g., clofibrate) and may result from increased activity of hydroxymethylglutaryl-coenzyme A (HMG-CoA) reductase, the rate-limiting enzyme of hepatic cholesterol synthesis, and increased hepatic uptake of cholesterol from blood. In patients with gallstones, dietary cholesterol increases biliary cholesterol secretion. This does not occur in non-gallstone patients on high-cholesterol diets. In addition to environmental factors such as high-caloric and cholesterol-rich diets, genetic factors play an important role in gallstone disease. A large study of symptomatic gallstones in Swedish twins provided strong evidence for a role of genetic factors in gallstone pathogenesis. Genetic factors accounted for 25%, shared environmental factors for 13%, and individual environmental factors for 62% of the phenotypic variation among monozygotic twins. A single nucleotide polymorphism of the gene encoding the hepatic cholesterol transporter ABCG5/G8 has been found in 21% of patients with gallstones, but only in 9% of the general population. It is thought to cause a gain of function of the cholesterol transporter and to contribute to cholesterol hypersecretion. A high prevalence of gallstones is found among first-degree relatives of gallstone carriers and in certain ethnic populations such as American Indians, Chilean Indians, and Chilean Hispanics. A common genetic trait has been identified for some of these populations by mitochondrial DNA analysis. In some patients, impaired hepatic conversion of cholesterol to bile acids may also occur, resulting in an increase of the lithogenic cholesterol/bile acid ratio. Although most cholesterol stones have a polygenic basis, there are rare monogenic (Mendelian) causes. Mutations in the *CYP7A1* gene have been described that result in a deficiency of the enzyme cholesterol 7-hydroxylase, which catalyzes

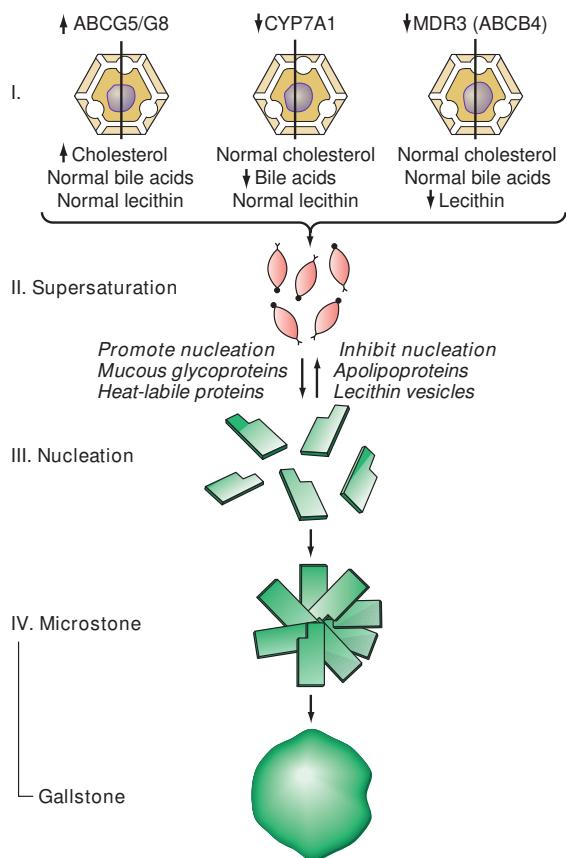


FIGURE 346-1 Scheme showing pathogenesis of cholesterol gallstone formation. Conditions or factors that increase the ratio of cholesterol to bile acids and phospholipids (lecithin) favor gallstone formation. ABCB4, ATP-binding cassette transporter; ABCG5/G8, ATP-binding cassette (ABC) transporter G5/G8; CYP7A1, cytochrome P450 7A1; MDR3, multidrug resistance protein 3, also called phospholipid export pump.

the initial step in cholesterol catabolism and bile acid synthesis. The homozygous state is associated with hypercholesterolemia and gallstones. Because the phenotype is expressed in the heterozygote state, mutations in the *CYP7A1* gene may contribute to the susceptibility to cholesterol gallstone disease in the population. Mutations in the *MDR3* (ABCB4) gene, which encodes the phospholipid export pump in the canalicular membrane of the hepatocyte, may cause defective phospholipid secretion into bile, resulting in cholesterol supersaturation of bile and formation of cholesterol gallstones in the gallbladder and in the bile ducts. Thus, an excess of biliary cholesterol in relation to bile acids and phospholipids is primarily due to hypersecretion of cholesterol, but hyposecretion of bile acids or phospholipids may contribute. An additional disturbance of bile acid metabolism that is likely to contribute to supersaturation of bile with cholesterol is enhanced conversion of cholic acid to deoxycholic acid, with replacement of the cholic acid pool by an expanded deoxycholic acid pool. It may result from enhanced dehydroxylation of cholic acid and increased absorption of newly formed deoxycholic acid. An increased deoxycholate secretion is associated with hypersecretion of cholesterol into bile.

While supersaturation of bile with cholesterol is an important prerequisite for gallstone formation, it is generally not sufficient by itself to produce cholesterol precipitation *in vivo*. Most individuals with supersaturated bile do not develop stones because the time required for cholesterol crystals to nucleate and grow is longer than the time bile remains in the gallbladder.

An important mechanism is *nucleation* of cholesterol monohydrate crystals, which is greatly accelerated in human lithogenic bile. Accelerated nucleation of cholesterol monohydrate in bile may be due to either

an excess of pronucleating factors or a deficiency of antinucleating factors. Mucin and certain nonmucin glycoproteins, principally immunoglobulins, appear to be pronucleating factors, while apolipoproteins A-I and A-II and other glycoproteins appear to be antinucleating factors. Pigment particles may possibly play a role as nucleating factors. In a genome-wide analysis of serum bilirubin levels, the uridine diphosphate-glucuronyltransferase 1A1 (*UGT1A1*) Gilbert's syndrome gene variant was associated with the presence of gallstone disease. Because most gallstones associated with the *UGT1A1* variant were cholesterol stones, this finding points to the role of pigment particles in the pathogenesis of gallbladder stones. Cholesterol monohydrate crystal nucleation and crystal growth probably occur within the mucin gel layer. Vesicle fusion leads to liquid crystals, which, in turn, nucleate into solid cholesterol monohydrate crystals. Continued growth of the crystals occurs by direct nucleation of cholesterol molecules from supersaturated unilamellar or multilamellar biliary vesicles.

A third important mechanism in cholesterol gallstone formation is *gallbladder hypomotility*. If the gallbladder emptied all supersaturated or crystal-containing bile completely, stones would not be able to grow. A high percentage of patients with gallstones exhibits abnormalities of gallbladder emptying. Ultrasonographic studies show that gallstone patients display an increased gallbladder volume during fasting and after a test meal (residual volume) and that fractional emptying after gallbladder stimulation is decreased. The incidence of gallstones is increased in conditions associated with infrequent or impaired gallbladder emptying such as fasting, parenteral nutrition, or pregnancy and in patients using drugs that inhibit gallbladder motility.

Biliary sludge is a thick, mucous material that, upon microscopic examination, reveals lecithin-cholesterol liquid crystals, cholesterol monohydrate crystals, calcium bilirubinate, and mucin gels. Biliary sludge typically forms a crescent-like layer in the most dependent portion of the gallbladder and is recognized by characteristic echoes on ultrasonography (see below). The presence of biliary sludge implies two abnormalities: (1) the normal balance between gallbladder mucin secretion and elimination has become deranged, and (2) nucleation of biliary solutes has occurred. That biliary sludge may be a precursor form of gallstone disease is evident from several observations. In one study, 96 patients with gallbladder sludge were followed prospectively by serial ultrasound studies. In 18%, biliary sludge disappeared and did not recur for at least 2 years. In 60%, biliary sludge disappeared and reappeared; in 14%, gallstones (8% asymptomatic, 6% symptomatic) developed; and in 6%, severe biliary pain with or without acute pancreatitis occurred. In 12 patients, cholecystectomies were performed, 6 for gallstone-associated biliary pain and 3 in symptomatic patients with sludge but without gallstones who had prior attacks of pancreatitis; the latter did not recur after cholecystectomy. It should be emphasized that biliary sludge can develop with disorders that cause gallbladder hypomotility; that is, surgery, burns, total parenteral nutrition, pregnancy, and oral contraceptives—all of which are associated with gallstone formation. However, the presence of biliary sludge implies supersaturation of bile with either cholesterol or calcium bilirubinate.

Two other conditions are associated with cholesterol-stone or biliary-sludge formation: pregnancy and rapid weight reduction through a very-low-calorie diet. There appear to be two key changes during pregnancy that contribute to a “cholelithogenic state”: (1) a marked increase in cholesterol saturation of bile during the third trimester and (2) sluggish gallbladder contraction in response to a standard meal, resulting in impaired gallbladder emptying. That these changes are related to pregnancy per se is supported by several studies that show reversal of these abnormalities quite rapidly after delivery. During pregnancy, gallbladder sludge develops in 20–30% of women and gallstones in 5–12%. Although biliary sludge is a common finding during pregnancy, it is usually asymptomatic and often resolves spontaneously after delivery. Gallstones, which are less common than sludge and frequently associated with biliary colic, may also disappear after delivery because of spontaneous dissolution related to bile becoming unsaturated with cholesterol postpartum.

Approximately 10–20% of persons with rapid weight reduction achieved through very-low-calorie dieting develop gallstones. In a

study involving 600 patients who completed a 3-month, 520-kcal/d diet, UDCA in a dosage of 600 mg/d proved highly effective in preventing gallstone formation; gallstones developed in only 3% of UDCA recipients, compared to 28% of placebo-treated patients. In obese patients treated by gastric banding, 500 mg/d of UDCA reduced the risk of gallstone formation from 30 to 8% within a follow-up of 6 months.

To summarize, cholesterol gallstone disease occurs because of several defects, which include (1) bile supersaturation with cholesterol, (2) nucleation of cholesterol monohydrate with subsequent crystal retention and stone growth, and (3) abnormal gallbladder motor function with delayed emptying and stasis. Other important factors known to predispose to cholesterol-stone formation are summarized in Table 346-1.

PIGMENT STONES Black pigment stones are composed of either pure calcium bilirubinate or polymer-like complexes with calcium and mucin glycoproteins. They are more common in patients who have chronic hemolytic states (with increased conjugated bilirubin in bile); cirrhosis, especially related to alcohol; Gilbert's syndrome; or cystic fibrosis. Gallbladder stones in patients with ileal diseases, ileal resection, or ileal bypass generally are also black pigment stones.

TABLE 346-1 Predisposing Factors for Cholesterol and Pigment Gallstone Formation

Cholesterol Stones

- Demographic/genetic factors: Prevalence highest in North American Indians, Chilean Indians, and Chilean Hispanics, greater in Northern Europe and North America than in Asia, lowest in Japan; familial disposition; hereditary aspects
- Obesity, metabolic syndrome: Normal bile acid pool and secretion but increased biliary secretion of cholesterol
- Rapid weight loss: Mobilization of tissue cholesterol leads to increased biliary cholesterol secretion while enterohepatic circulation of bile acids is decreased
- Female sex hormones
 - Estrogens stimulate hepatic lipoprotein receptors, increase uptake of dietary cholesterol, and increase biliary cholesterol secretion
 - Natural estrogens, other estrogens, and oral contraceptives lead to decreased bile salt secretion and decreased conversion of cholesterol to cholesterol esters
- Pregnancy: Impaired gallbladder emptying caused by progesterone combined with the influence of estrogens, which increase biliary cholesterol secretion
- Increasing age: Increased biliary secretion of cholesterol, decreased size of bile acid pool, decreased secretion of bile salts
- Gallbladder hypomotility leading to stasis and formation of sludge
 - Total parenteral nutrition
 - Fasting
 - Pregnancy
 - Drugs such as octreotide
- Clofibrate therapy: Increased biliary secretion of cholesterol
- Decreased bile acid secretion
 - Genetic defect of the *CYP7A1* gene
- Decreased phospholipid secretion: Genetic defect of the *MDR3* gene
- Miscellaneous
 - High-calorie, high-fat diet
 - Spinal cord injury

Pigment Stones

- Demographic/genetic factors: Asia, rural setting (presumed due to increased prevalence of parasitic biliary infections; the incidence has been dropping with time)
- Chronic hemolysis (example: sickle cell disease)
- Alcohol related liver cirrhosis
- Ineffective erythropoiesis (example: pernicious anemia)
- Cystic fibrosis
- Chronic biliary tract infection, parasite infections
- Increasing age
- Ileal disease, ileal resection or bypass

Enterohepatic recycling of bilirubin in ileal disease states contributes to their pathogenesis. Brown pigment stones are composed of calcium salts of unconjugated bilirubin with varying amounts of cholesterol and protein. They are caused by the presence of increased amounts of unconjugated, insoluble bilirubin in bile that precipitates to form stones. Deconjugation of an excess of soluble bilirubin mono- and diglucuronides may be mediated by endogenous β -glucuronidase but may also occur by spontaneous hydrolysis. Sometimes, the enzyme is also produced when bile is chronically infected by bacteria, and such stones are brown. Pigment stone formation is frequent in Asia and is often associated with parasitic infections in the gallbladder and biliary tree (Table 346-1).

Diagnosis Procedures of potential use in the diagnosis of cholelithiasis and other diseases of the gallbladder are detailed in Table 346-2. Ultrasonography of the gallbladder is very accurate in the identification of cholelithiasis and has replaced oral cholecystography (OCG) (Fig. 346-2A). Stones as small as 1.5 mm in diameter may be confidently identified provided that firm criteria are used (e.g., acoustic “shadowing” of opacities that are within the gallbladder lumen and that change with the patient’s position [by gravity]). In major medical centers, the false-negative and false-positive rates for ultrasound in gallstone patients are ~2–4%. Biliary sludge is material of low echogenic activity that typically forms a layer in the most dependent position of the gallbladder. This layer shifts with postural changes but fails to produce acoustic shadowing; these two characteristics distinguish sludges from gallstones. Ultrasound can also be used to assess the emptying function of the gallbladder.

TABLE 346-2 Diagnostic Evaluation of the Gallbladder

DIAGNOSTIC ADVANTAGES	DIAGNOSTIC LIMITATIONS	COMMENT
Ultrasound		
Rapid Accurate identification of gallstones (>95%) Simultaneous scanning of GB, liver, bile ducts, pancreas “Real-time” scanning allows assessment of GB volume, contractility Not limited by jaundice, pregnancy May detect very small stones	Bowel gas Massive obesity Ascites	Procedure of choice for detection of stones
Plain Abdominal X-Ray		
Low cost Readily available	Relatively low yield Contraindicated in pregnancy	Pathognomonic findings in: calcified gallstones, limey bile, porcelain GB, emphysematous cholecystitis, gallstone ileus
Cholescintigraphy (HIDA, DISIDA, etc.)		
Accurate identification of cystic duct obstruction Simultaneous assessment of bile ducts	Contraindicated in pregnancy Serum bilirubin >103–205 $\mu\text{mol/L}$ (6–12 mg/dL) Cholecystogram of low resolution	Indicated for confirmation of suspected acute cholecystitis; less sensitive and less specific in chronic cholecystitis; useful in the diagnosis of acalculous cholecystopathy, especially if given with CCK to assess GB emptying

Abbreviations: CCK, cholecystokinin; DISIDA, diisopropyl iminodiacetic acid; GB, gallbladder; HIDA, hydroxyl iminodiacetic acid.

The plain abdominal film may detect gallstones containing sufficient calcium to be radiopaque (10–15% of cholesterol and ~50% of pigment stones). Plain radiography may also be of use in the diagnosis of emphysematous cholecystitis, porcelain gallbladder, limey bile, and gallstone ileus.

OCG was historically a useful procedure for the diagnosis of gallstones but has been replaced by ultrasound and is now regarded as obsolete. It may be used to assess the patency of the cystic duct and gallbladder emptying function. Further, OCG can also delineate the size and number of gallstones and determine whether they are calcified, useful information if medical dissolution is being considered.

Radiopharmaceuticals such as $^{99\text{m}}\text{Tc}$ -labeled *N*-substituted iminodiacetic acids (HIDA and DISIDA) are rapidly extracted from the blood and are excreted into the biliary tree in high concentration even in the presence of mild to moderate serum bilirubin elevations. Failure to image the gallbladder in the presence of biliary ductal visualization may indicate cystic duct obstruction, acute or chronic cholecystitis, or surgical absence of the organ. Such scans have application in the diagnosis of acute cholecystitis and may play a role in the detection of a postcholecystectomy bile leak.

Symptoms of Gallstone Disease Gallstones usually produce symptoms by causing inflammation or obstruction following their migration into the cystic duct or CBD. The most specific and characteristic symptom of gallstone disease is biliary colic that is a constant and often long-lasting pain (see below). Obstruction of the cystic duct or CBD by a stone produces increased intraluminal pressure and distention of the viscus that cannot be relieved by repetitive biliary contractions. The resultant visceral pain is characteristically a severe, steady ache or fullness in the epigastrium or right upper quadrant (RUQ) of the abdomen with frequent radiation to the interscapular area, right scapula, or shoulder.

Biliary colic begins quite suddenly and may persist with severe intensity for 30 min to 5 h, subsiding gradually or rapidly. It is steady rather than intermittent, as would be suggested by the word *colic*, which must be regarded as a misnomer, although it is in widespread use. An episode of biliary pain persisting beyond 5 h should raise the suspicion of acute cholecystitis (see below). Nausea and vomiting frequently accompany episodes of biliary pain. An elevated level of serum bilirubin and/or alkaline phosphatase suggests a common duct stone. Fever or chills (rigors) with biliary pain usually imply a complication, that is, cholecystitis, pancreatitis, or cholangitis. Complaints of short-lasting, vague epigastric fullness, dyspepsia, eructation, or flatulence, especially following a fatty meal, should not be confused with biliary pain. Such symptoms are frequently elicited from patients with or without gallstone disease but are not specific for biliary calculi. Biliary colic may be precipitated by eating a fatty meal, by consumption of a large meal following a period of prolonged fasting, or by eating a normal meal; it is frequently nocturnal, occurring within a few hours of retiring.

Natural History Gallstone disease discovered in an asymptomatic patient or in a patient whose symptoms are not referable to cholelithiasis is a common clinical problem. Sixty to 80% of persons with asymptomatic gallstones remain asymptomatic over follow-up periods of up to 25 years. The probability of developing symptoms within 5 years after diagnosis is 2–4% per year and decreases in the years thereafter to 1–2%. The yearly incidence of complications is about 0.1–0.3%. Patients remaining asymptomatic for 15 years were found to be unlikely to develop symptoms during further follow-up, and most patients who did develop complications from their gallstones experienced *prior* warning symptoms. Similar conclusions apply to diabetic patients with silent gallstones. Decision analysis has suggested that (1) the cumulative risk of death due to gallstone disease while on expectant management is small, and (2) prophylactic cholecystectomy is not warranted.

Complications requiring cholecystectomy are much more common in gallstone patients who have developed symptoms of biliary pain. Patients found to have gallstones at a young age are more likely to develop symptoms from cholelithiasis than are patients >60 years at the time of initial diagnosis. Patients with diabetes mellitus and gallstones

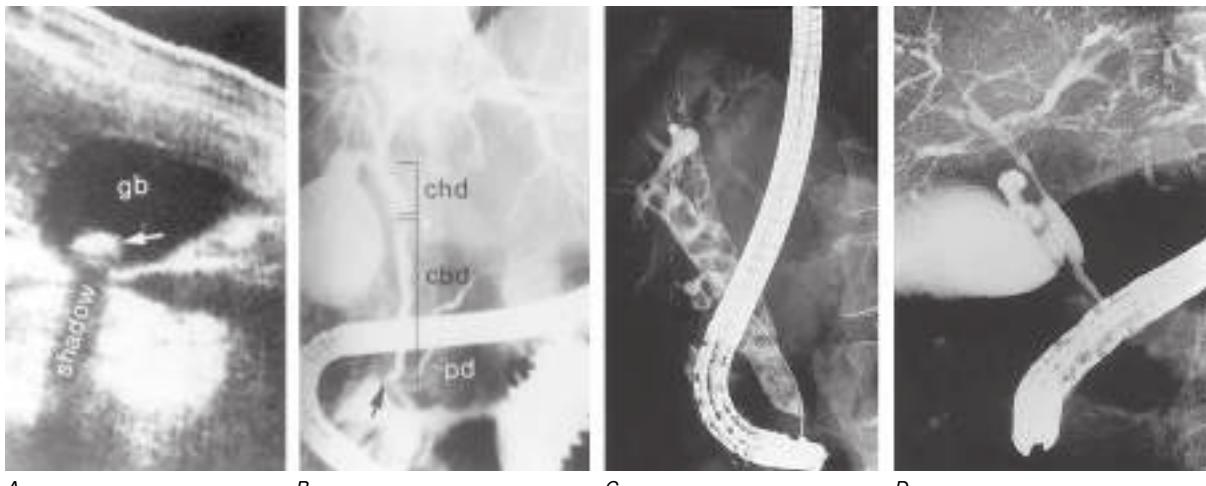


FIGURE 346-2 Examples of ultrasound and radiologic studies of the biliary tract. *A*. An ultrasound study showing a distended gallbladder (GB) containing a single large stone (arrow), which casts an acoustic shadow. *B*. Endoscopic retrograde cholangiopancreatogram (ERCP) showing normal biliary tract anatomy. In addition to the endoscope and large vertical gallbladder filled with contrast dye, the common hepatic duct (CHD), common bile duct (CBD), and pancreatic duct (PD) are shown. The arrow points to the ampulla of Vater. *C*. Endoscopic retrograde cholangiogram (ERC) showing choledocholithiasis. The biliary tract is dilated and contains multiple radiolucent calculi. *D*. ERCP showing sclerosing cholangitis. The CBD shows areas that are strictured and narrowed.

may be somewhat more susceptible to septic complications, but the magnitude of risk of septic biliary complications in diabetic patients is incompletely defined.

TREATMENT

Gallstones

SURGICAL THERAPY

In asymptomatic gallstone patients, the risk of developing symptoms or complications requiring surgery is quite small (see above). Thus, a recommendation for cholecystectomy in a patient with gallstones should probably be based on assessment of three factors: (1) the presence of symptoms that are frequent enough or severe enough to interfere with the patient's general routine; (2) the presence of a prior complication of gallstone disease, that is, history of acute cholecystitis, pancreatitis, gallstone fistula, etc.; or (3) the presence of an underlying condition predisposing the patient to increased risk of gallstone complications (e.g., a previous attack of acute cholecystitis regardless of current symptomatic status). Patients with very large gallstones (>3 cm in diameter) and patients harboring gallstones in a congenitally anomalous gallbladder might also be considered for prophylactic cholecystectomy. Although young age is a worrisome factor in asymptomatic gallstone patients, few authorities would now recommend routine cholecystectomy in young patients with silent stones. Laparoscopic cholecystectomy is a minimal-access approach for the removal of the gallbladder together with its stones. Its advantages include a markedly shortened hospital stay, minimal disability, and decreased cost, and it is the procedure of choice for most patients referred for elective cholecystectomy.

From several studies involving >4000 patients undergoing laparoscopic cholecystectomy, the following key points emerge: (1) complications develop in $\sim 4\%$ of patients, (2) conversion to laparotomy occurs in 5%, (3) the death rate is remarkably low (i.e., $<0.1\%$), and (4) the rate of bile duct injuries is low (i.e., 0.2–0.6%) and comparable with open cholecystectomy. These data indicate why laparoscopic cholecystectomy has become the "gold standard" for treating symptomatic cholelithiasis.

MEDICAL THERAPY—GALLSTONE DISSOLUTION

In carefully selected patients with a functioning gallbladder and with radiolucent stones <10 mm in diameter, complete dissolution can be achieved in $\sim 50\%$ of patients within 6 months to 2 years. For

good results within a reasonable time period, this therapy should be limited to radiolucent stones <5 mm in diameter. The dose of UDCA should be 10–15 mg/kg per d. Stones >10 mm in size rarely dissolve. Pigment stones are not responsive to UDCA therapy. Probably $\leq 10\%$ of patients with *symptomatic* cholelithiasis are candidates for such treatment. However, in addition to the vexing problem of recurrent stones (30–50% over 3–5 years of follow-up), there is also the factor of taking a drug for up to 2 years and perhaps indefinitely. The advantages and success of laparoscopic cholecystectomy have largely reduced the role of gallstone dissolution to patients who wish to avoid or are not candidates for elective cholecystectomy. However, patients with cholesterol gallstone disease who develop recurrent choledocholithiasis after cholecystectomy should be on long-term treatment with UDCA.

■ ACUTE AND CHRONIC CHOLECYSTITIS

Acute Cholecystitis Acute inflammation of the gallbladder wall usually follows obstruction of the cystic duct by a stone. Inflammatory response can be evoked by three factors: (1) *mechanical inflammation* produced by increased intraluminal pressure and distention with resulting ischemia of the gallbladder mucosa and wall, (2) *chemical inflammation* caused by the release of lysophosphatidylcholine (due to the action of phospholipase on lecithin in bile) and other local tissue factors, and (3) *bacterial inflammation*, which may play a role in 50–85% of patients with acute cholecystitis. The organisms most frequently isolated by culture of gallbladder bile in these patients include *Escherichia coli*, *Klebsiella* spp., *Streptococcus* spp., and *Clostridium* spp.

Acute cholecystitis often begins as an attack of biliary pain that progressively worsens. Approximately 60–70% of patients report having experienced prior attacks that resolved spontaneously. As the episode progresses, however, the pain of acute cholecystitis becomes more generalized in the right upper abdomen. As with biliary colic, the pain of cholecystitis may radiate to the interscapular area, right scapula, or shoulder. Peritoneal signs of inflammation such as increased pain with jarring or on deep respiration may be apparent. The patient is anorectic and often nauseated. Vomiting is relatively common and may produce symptoms and signs of vascular and extracellular volume depletion. Jaundice is unusual early in the course of acute cholecystitis but may occur when edematous inflammatory changes involve the bile ducts and surrounding lymph nodes.

A low-grade fever is characteristically present, but shaking chills or rigors are not uncommon. The RUQ of the abdomen is almost

invariably tender to palpation. An enlarged, tense gallbladder is palpable in 25–50% of patients. Deep inspiration or cough during subcostal palpation of the RUQ usually produces increased pain and inspiratory arrest (Murphy's sign). Localized rebound tenderness in the RUQ is common, as are abdominal distention and hypoactive bowel sounds from paralytic ileus, but generalized peritoneal signs and abdominal rigidity are usually lacking, in the absence of perforation.

The diagnosis of acute cholecystitis is usually made on the basis of a characteristic history and physical examination. The triad of sudden onset of RUQ tenderness, fever, and leukocytosis is highly suggestive. Typically, leukocytosis in the range of 10,000–15,000 cells per microliter with a left shift on differential count is found. The serum bilirubin is mildly elevated (>85.5 μmol/L [5 mg/dL]) in fewer than half of patients, whereas about one-fourth have modest elevations in serum aminotransferases (usually less than a fivefold elevation). Ultrasound will demonstrate calculi in 90–95% of cases and is useful for detection of signs of gallbladder inflammation including thickening of the wall, pericholecystic fluid, and dilatation of the bile duct. The radionuclide (e.g., HIDA) biliary scan may be confirmatory if bile duct imaging is seen without visualization of the gallbladder.

Approximately 75% of patients treated medically have remission of acute symptoms within 2–7 days following hospitalization. In 25%, however, a complication of acute cholecystitis will occur despite conservative treatment (see below). In this setting, prompt surgical intervention is required. Of the 75% of patients with acute cholecystitis who undergo remission of symptoms, ~25% will experience a recurrence of cholecystitis within 1 year, and 60% will have at least one recurrent bout within 6 years. In view of the natural history of the disease, acute cholecystitis is best treated by early surgery whenever possible. Mirizzi's syndrome is a rare complication in which a gallstone becomes impacted in the cystic duct or neck of the gallbladder causing compression of the CBD, resulting in CBD obstruction and jaundice. Ultrasound shows gallstone(s) lying outside the hepatic duct. Endoscopic retrograde cholangiopancreatography (ERCP) (Fig. 346-2B), percutaneous transhepatic cholangiography (PTC), or magnetic resonance cholangio-pancreatography (MRCP) will usually demonstrate the characteristic extrinsic compression of the CBD. Surgery consists of removing the cystic duct, diseased gallbladder, and impacted stone. The preoperative diagnosis of Mirizzi's syndrome is important to avoid CBD injury.

ACALCULOUS CHOLECYSTITIS In 5–14% of patients with acute cholecystitis, calculi obstructing the cystic duct are not found at surgery. In >50% of such cases, an underlying explanation for acalculous inflammation is not found. An increased risk for the development of acalculous cholecystitis is especially associated with prolonged fasting, with serious trauma or burns, in the postpartum period following prolonged labor, and with orthopedic and other nonbiliary major surgical operations in the postoperative period. It may possibly complicate periods of prolonged parenteral hyperalimentation. For some of these cases, biliary sludge in the cystic duct may be responsible. Other precipitating factors include vasculitis, obstructing adenocarcinoma of the gallbladder, diabetes mellitus, torsion of the gallbladder, "unusual" bacterial infections of the gallbladder (e.g., *Leptospira*, *Streptococcus*, *Salmonella*, or *Vibrio cholerae*), and parasitic infestation of the gallbladder. Acalculous cholecystitis may also be seen with a variety of other systemic disease processes (e.g., sarcoidosis, cardiovascular disease, tuberculosis, syphilis, actinomycosis).

Although the clinical manifestations of acalculous cholecystitis are indistinguishable from those of calculous cholecystitis, the setting of acute gallbladder inflammation complicating severe underlying illness is characteristic of acalculous disease. Ultrasound or computed tomography (CT) examinations demonstrating a large, tense, static gallbladder without stones and with evidence of poor emptying over a prolonged period may be diagnostically useful in some cases. The complication rate for acalculous cholecystitis exceeds that for calculous cholecystitis. Successful management of acute acalculous cholecystitis appears to depend primarily on early diagnosis and surgical intervention, with meticulous attention to postoperative care.

ACALCULOUS CHOLECYSTOPATHY Disordered motility of the gallbladder can produce recurrent biliary pain in patients without gallstones. Infusion of an octapeptide of CCK can be used to measure the gallbladder ejection fraction during cholescintigraphy. The surgical findings have included abnormalities such as chronic cholecystitis, gallbladder muscle hypertrophy, and/or a markedly narrowed cystic duct. Some of these patients may well have had antecedent gallbladder disease. The following criteria can be used to identify patients with acalculous cholecystopathy: (1) recurrent episodes of typical RUQ pain characteristic of biliary tract pain, (2) abnormal CCK cholescintigraphy demonstrating a gallbladder ejection fraction of <40%, and (3) infusion of CCK reproducing the patient's pain. An additional clue would be the identification of a large gallbladder on ultrasound examination. Importantly, it should be noted that SOD dysfunction can also give rise to recurrent RUQ pain and CCK-scintigraphic abnormalities.

EMPHYSEMATOUS CHOLECYSTITIS So-called emphysematous cholecystitis is thought to begin with acute cholecystitis (calculous or acalculous) followed by ischemia or gangrene of the gallbladder wall and infection by gas-producing organisms. Bacteria most frequently cultured in this setting include anaerobes, such as *Clostridium welchii* or *C. perfringens*, and aerobes, such as *E. coli*. This condition occurs most frequently in elderly men and in patients with diabetes mellitus. The clinical manifestations are essentially indistinguishable from those of nongaseous cholecystitis. The diagnosis is usually made on plain abdominal film by finding gas within the gallbladder lumen, dissecting within the gallbladder wall to form a gaseous ring, or in the pericholecystic tissues. The morbidity and mortality rates with emphysematous cholecystitis are considerable. Prompt surgical intervention coupled with appropriate antibiotics is mandatory.

Chronic Cholecystitis Chronic inflammation of the gallbladder wall is almost always associated with the presence of gallstones and is thought to result from repeated bouts of subacute or acute cholecystitis or from persistent mechanical irritation of the gallbladder wall by gallstones. The presence of bacteria in the bile occurs in >25% of patients with chronic cholecystitis. The presence of infected bile in a patient with chronic cholecystitis undergoing elective cholecystectomy probably adds little to the operative risk. Chronic cholecystitis may be asymptomatic for years, which may progress to symptomatic gallbladder disease or to acute cholecystitis or may present with complications (see below).

Complications of Cholecystitis • EMPYEMA AND HYDROPS Empyema of the gallbladder usually results from progression of acute cholecystitis with persistent cystic duct obstruction to superinfection of the stagnant bile with a pus-forming bacterial organism. The clinical picture resembles that of cholangitis with high fever; severe RUQ pain; marked leukocytosis; and often, prostration. Empyema of the gallbladder carries a high risk of gram-negative sepsis and/or perforation. Emergency surgical intervention with proper antibiotic coverage is required as soon as the diagnosis is suspected.

Hydrops or mucocele of the gallbladder may also result from prolonged obstruction of the cystic duct, usually by a large solitary calculus. In this instance, the obstructed gallbladder lumen is progressively distended, over a period of time, by mucus (mucocele) or by a clear transudate (hydrops) produced by mucosal epithelial cells. A visible, easily palpable, non-tender mass sometimes extending from the RUQ into the right iliac fossa may be found on physical examination. The patient with hydrops of the gallbladder frequently remains asymptomatic, although chronic RUQ pain may also occur. Cholecystectomy is indicated, because empyema, perforation, or gangrene may complicate the condition.

GANGRENE AND PERFORATION Gangrene of the gallbladder results from ischemia of the wall and patchy or complete tissue necrosis. Underlying conditions often include marked distention of the gallbladder, vasculitis, diabetes mellitus, empyema, or torsion resulting in arterial occlusion. Gangrene usually predisposes to perforation of the gallbladder, but perforation may also occur in chronic cholecystitis without premonitory warning symptoms. *Localized perforations*

are usually contained by the omentum or by adhesions produced by recurrent inflammation of the gallbladder. Bacterial superinfection of the walled-off gallbladder contents results in abscess formation. Most patients are best treated with cholecystectomy, but some seriously ill patients may be managed with cholecystostomy and drainage of the abscess. *Free perforation* is less common but is associated with a mortality rate of ~30%. Such patients may experience a sudden transient relief of RUQ pain as the distended gallbladder decompresses; this is followed by signs of generalized peritonitis.

FISTULA FORMATION AND GALLSTONE ILEUS Fistula formation into an adjacent organ adherent to the gallbladder wall may result from inflammation and adhesion formation. Fistulas into the duodenum are most common, followed in frequency by those involving the hepatic flexure of the colon, stomach or jejunum, abdominal wall, and renal pelvis. Clinically “silent” biliary-enteric fistulas occurring as a complication of acute cholecystitis have been found in up to 5% of patients undergoing cholecystectomy. Asymptomatic cholecystoenteric fistulas may sometimes be diagnosed by finding gas in the biliary tree on plain abdominal films. Barium contrast studies or endoscopy of the upper gastrointestinal tract or colon may demonstrate the fistula. Treatment in the symptomatic patient usually consists of cholecystectomy, CBD exploration, and closure of the fistulous tract.

Gallstone ileus refers to mechanical intestinal obstruction resulting from the passage of a large gallstone into the bowel lumen. The stone customarily enters the duodenum through a cholecystoenteric fistula at that level. The site of obstruction by the impacted gallstone is usually at the ileocecal valve, provided that the more proximal small bowel is of normal caliber. Most patients do not give a history of either prior biliary tract symptoms or complaints suggestive of acute cholecystitis or fistula formation. Large stones, >2.5 cm in diameter, are thought to predispose to fistula formation by gradual erosion through the gallbladder fundus. Diagnostic confirmation may occasionally be found on the plain abdominal film (e.g., small-intestinal obstruction with gas in the biliary tree [pneumobilia] and a calcified, ectopic gallstone) or following an upper gastrointestinal series (cholecystoduodenal fistula with small-bowel obstruction at the ileocecal valve). Laparotomy with stone extraction (or propulsion into the colon) remains the procedure of choice to relieve obstruction. Evacuation of large stones within the gallbladder should also be performed. In general, the gallbladder and its attachment to the intestines should be left alone.

LIMEY MILK OF CALCIUM BILE AND PORCELAIN GALLBLADDER Calcium salts in the lumen of the gallbladder in sufficient concentration may produce calcium precipitation and diffuse, hazy opacification of bile or a layering effect on plain abdominal roentgenography. This so-called limey bile, or milk of calcium bile, is usually clinically innocuous, but cholecystectomy is often performed, especially when it occurs in a hydropic gallbladder. In the entity called *porcelain gallbladder*, calcium salt deposition within the wall of a chronically inflamed gallbladder may be detected on the plain abdominal film. In the past, cholecystectomy was advised in all patients with porcelain gallbladder because there was felt to be a high incidence of carcinoma of the gallbladder associated with this condition, an association challenged by a number of studies. Two patterns of gallbladder wall calcification have now been appreciated: complete intramural calcification and selective mucosal calcification. The incidence of cancer in those with selective intramural calcification is higher than those with complete mucosal wall calcification, but the risk is very small. As such, the need for cholecystectomy for porcelain gallbladder is not absolute; close surveillance in these patients is also acceptable.

TREATMENT

Acute Cholecystitis

MEDICAL THERAPY

Although surgical intervention remains the mainstay of therapy for acute cholecystitis and its complications, a period of in-hospital stabilization may be required before cholecystectomy. Oral intake is

eliminated, nasogastric suction may be indicated, extracellular volume depletion and electrolyte abnormalities are repaired, and analgesia is provided. Intravenous antibiotic therapy is indicated in patients with severe acute cholecystitis, even though bacterial superinfection of bile may not have occurred in the early stages of the inflammatory process. Antibiotic therapy is guided by the most common organisms likely to be present including *E. coli*, *Klebsiella*, *Enterococcus*, *Enterobacter*, and *Streptococcus*. Effective antibiotics include piperacillin plus tazobactam, imipenem, meropenem, ceftriaxone plus metronidazole, and levofloxacin plus metronidazole (Chap. 161). Postoperative complications of wound infection, abscess formation, and sepsis are reduced in antibiotic-treated patients.

SURGICAL THERAPY

The optimal timing of surgical intervention in patients with acute cholecystitis depends on stabilization of the patient. The clear trend is toward earlier surgery, and this is due in part to requirements for shorter hospital stays. Urgent (emergency) cholecystectomy or percutaneous cholecystostomy is probably appropriate in most patients in whom a complication of acute cholecystitis such as empyema, emphysematous cholecystitis, or perforation is suspected or confirmed. Patients with uncomplicated acute cholecystitis should undergo early elective laparoscopic cholecystectomy, ideally within 48–72 h after diagnosis. The complication rate is not increased in patients undergoing early as opposed to delayed (>6 weeks after diagnosis) cholecystectomy. Delayed surgical intervention is probably best reserved for (1) patients in whom the overall medical condition imposes an unacceptable risk for early surgery and (2) patients in whom the diagnosis of acute cholecystitis is in doubt. Thus, early cholecystectomy (within 72 h) is the treatment of choice for most patients with acute cholecystitis. Mortality figures for emergency cholecystectomy in most centers range from 1 to 3%, whereas the mortality risk for early elective cholecystectomy is ~0.5% in patients under age 60. Of course, the operative risks increase with age-related diseases of other organ systems and with the presence of long- or short-term complications of gallbladder disease. Seriously ill or debilitated patients with cholecystitis may be managed with percutaneous drainage (a cholecystostomy tube), transpapillary drainage (an endoscopically placed transpapillary drainage catheter via the cystic duct), or transmural drainage (an endoscopically placed covered, lumen-apposing stent). Elective cholecystectomy may then be done at a later date.

Postcholecystectomy Complications Early complications following cholecystectomy include atelectasis and other pulmonary disorders, abscess formation (often subphrenic), external or internal hemorrhage, biliary-enteric fistula, and bile leaks. Jaundice may indicate absorption of bile from an intraabdominal collection following a biliary leak or mechanical obstruction of the CBD by retained calculi, intraductal blood clots, or extrinsic compression.

Overall, cholecystectomy is a very successful operation that provides total or near-total relief of preoperative symptoms in 75–90% of patients. The most common cause of persistent postcholecystectomy symptoms is an overlooked symptomatic nonbiliary disorder (e.g., reflux esophagitis, peptic ulceration, pancreatitis, or—most often—irritable bowel syndrome). In a small percentage of patients, however, a disorder of the extrahepatic bile ducts may result in persistent symptomatology. These so-called postcholecystectomy syndromes may be due to (1) biliary strictures, (2) retained biliary calculi, (3) cystic duct stump syndrome, (4) stenosis or dyskinesia of the SOD, or (5) bile salt-induced diarrhea or gastritis.

CYSTIC DUCT STUMP SYNDROME In the absence of cholangiographically demonstrable retained stones, symptoms resembling biliary pain or cholecystitis in the postcholecystectomy patient have frequently been attributed to disease in a long (>1 cm) cystic duct remnant (cystic duct stump syndrome). Careful analysis, however, reveals that postcholecystectomy complaints are attributable to other causes in almost all patients in whom the symptom complex was originally thought to result from the existence of a long cystic duct stump.

Accordingly, considerable care should be taken to investigate the possible role of other factors in the production of postcholecystectomy symptoms before attributing them to cystic duct stump syndrome.

SOD STENOSIS, SOD DYSKINESIA, AND BILIARY DYSKINESIA Symptoms of biliary colic accompanied by signs of recurrent, intermittent biliary obstruction may be produced by acalculous cholecystopathy, SOD stenosis, or SOD dyskinesia. SOD stenosis is thought to result from acute or chronic inflammation of the papilla of Vater or from glandular hyperplasia of the papillary segment. Five criteria have been used to define SOD stenosis: (1) upper abdominal pain, usually RUQ or epigastric; (2) abnormal liver tests; (3) dilatation of the CBD upon MRCP or ERCP examination; (4) delayed (>45 min) drainage of contrast material from the duct; and (5) increased basal pressure of the SOD. After exclusion of acalculous cholecystopathy, treatment consists of endoscopic or surgical sphincteroplasty to ensure wide patency of the distal portions of both the bile and pancreatic ducts. The greater the number of the preceding criteria present, the greater is the likelihood that a patient does have a degree of SOD sufficient to justify correction. The factors usually considered as indications for sphincterotomy include (1) prolonged duration of symptoms, (2) lack of response to symptomatic treatment, (3) presence of severe disability, and (4) the patient's choice of sphincterotomy over surgery (given a clear understanding on his or her part of the risks involved in both procedures).

Biliary SOD disorders are characterized by three criteria: (1) biliary pain, (2) absence of bile duct stones or other abnormalities, and (3) elevated liver enzymes or a dilated CBD, but not both. In this setting, either hepatobiliary scintigraphy or SOD manometry can support the diagnosis. Importantly, the presence of both elevated liver enzymes and a dilated CBD should raise the question of obstruction. Proposed mechanisms to account for SOD dysfunction include spasm of the sphincter, denervation sensitivity resulting in hypertonicity, and abnormalities in the sequencing or frequency rates of the sphincteric-contraction waves. When thorough evaluation has failed to demonstrate another cause for the pain and when cholangiographic and manometric criteria suggest a diagnosis of SOD dyskinesia, medical treatment with nitrates or anticholinergics to attempt pharmacologic relaxation of SOD has been proposed but not evaluated in detailed studies. Endoscopic biliary sphincterotomy (EBS) or surgical sphincterotomy may be indicated in patients who fail to respond to a 2- to 3-month trial of medical therapy, especially if SOD pressures are elevated. Approximately 45% of such patients have long-term pain relief after EBS. EBS has become the procedure of choice for removing bile duct stones and for other biliary and pancreatic problems.

BILE SALT INDUCED DIARRHEA AND GASTRITIS Postcholecystectomy patients may develop symptoms of dyspepsia, which have been attributed to duodenogastric reflux of bile. However, firm data linking these symptoms to bile gastritis after surgical removal of the gallbladder are lacking. Cholecystectomy induces persistent changes in gut transit, and these changes effect a noticeable modification of bowel habits. Cholecystectomy shortens gut transit time by accelerating passage of the fecal bolus through the colon with marked acceleration in the right colon, thus causing an increase in colonic bile acid output and a shift in bile acid composition toward the more diarrheagenic secondary bile acids, that is, deoxycholic acid. Diarrhea that is severe enough, that is, three or more watery movements per day, can be classified as postcholecystectomy diarrhea, and this occurs in 5–10% of patients undergoing elective cholecystectomy. Treatment with bile acid–sequestering agents such as colestyramine or colestipol is often effective in ameliorating troublesome diarrhea.

■ THE HYPERPLASTIC CHOLECYSTOSES

The term *hyperplastic cholecystoses* is used to denote a group of disorders of the gallbladder characterized by excessive proliferation of normal tissue components.

Adenomyomatosis is characterized by a benign proliferation of gallbladder surface epithelium with glandlike formations, extramural sinuses, transverse strictures, and/or fundal nodule (“adenoma” or “adenomyoma”) formation.

Cholesterosis is characterized by abnormal deposition of lipid, especially cholesterol esters, within macrophages in the lamina propria of the gallbladder wall. In its diffuse form (“strawberry gallbladder”), the gallbladder mucosa is brick red and speckled with bright yellow flecks of lipid. The localized form shows solitary or multiple “cholesterol polyps” studding the gallbladder wall. Cholesterol stones of the gallbladder are found in nearly half the cases. Cholecystectomy is only indicated in adenomyomatosis or cholesterosis when biliary symptoms are present.

The prevalence of gallbladder polyps in the adult population is 1–4% with a marked male predominance. Types of gallbladder polyps include cholesterol polyps, adenomyomas, inflammatory polyps, and adenomas (rare). No significant changes have been found over a 5-year period in asymptomatic patients with gallbladder polyps <6 mm and few changes in polyps 7–9 mm. Cholecystectomy is recommended in symptomatic patients as well as in asymptomatic patients >50 years whose polyps are >10 mm or associated with gallstones or polyp growth on serial ultrasonography.

DISEASES OF THE BILE DUCTS

■ CONGENITAL ANOMALIES

Biliary Atresia and Hypoplasia Atretic and hypoplastic lesions of the extrahepatic and large intrahepatic bile ducts are the most common biliary anomalies of clinical relevance encountered in infancy. The clinical picture is one of severe obstructive jaundice during the first month of life, with pale stools. When biliary atresia is suspected on the basis of clinical, laboratory, and imaging findings, the diagnosis is confirmed by surgical exploration and operative cholangiography. Approximately 10% of cases of biliary atresia are treatable with Roux-en-Y choledochojejunostomy, with the Kasai procedure (hepatic portoenterostomy) being attempted in the remainder in an effort to restore some bile flow. Most patients, even those having successful biliary-enteric anastomoses, eventually develop chronic cholangitis, extensive hepatic fibrosis, and portal hypertension.

Choledochal Cysts Cystic dilatation may involve the free portion of the CBD, that is, choledochal cyst, or may present as diverticulum formation in the intraduodenal segment. In the latter situation, chronic reflux of pancreatic juice into the biliary tree can produce inflammation and stenosis of the extrahepatic bile ducts, leading to cholangitis or biliary obstruction. Because the process may be gradual, ~50% of patients present with onset of symptoms after age 10. The diagnosis may be made by ultrasound, abdominal CT, MRCP, or cholangiography. Only one-third of patients show the classic triad of abdominal pain, jaundice, and an abdominal mass. Ultrasonographic detection of a cyst separate from the gallbladder should suggest the diagnosis of choledochal cyst, which can be confirmed by demonstrating the entrance of extrahepatic bile ducts into the cyst. Surgical treatment involves excision of the “cyst” and biliary-enteric anastomosis. Patients with choledochal cysts are at increased risk for the subsequent development of cholangiocarcinoma.

Congenital Biliary Ectasia Dilatation of intrahepatic bile ducts may involve either the major intrahepatic radicles (Caroli's disease), the inter- and intralobular ducts (congenital hepatic fibrosis), or both. In Caroli's disease, clinical manifestations include recurrent cholangitis, abscess formation in and around the affected ducts, and, often, brown pigment gallstone formation within portions of ectatic intrahepatic biliary radicles. Ultrasound, MRCP, and CT are of great diagnostic value in demonstrating cystic dilatation of the intrahepatic bile ducts. Treatment with ongoing antibiotic therapy is usually undertaken in an effort to limit the frequency and severity of recurrent bouts of cholangitis. Progression to secondary biliary cirrhosis with portal hypertension, extrahepatic biliary obstruction, cholangiocarcinoma, or recurrent episodes of sepsis with hepatic abscess formation is common.

■ CHOLEDOCHOLITHIASIS

Pathophysiology and Clinical Manifestations Passage of gallstones into the CBD occurs in ~10–15% of patients with cholelithiasis.

2650 The incidence of common duct stones increases with increasing age of the patient, so up to 25% of elderly patients may have calculi in the common duct at the time of cholecystectomy. Undetected duct stones are left behind in ~1–5% of cholecystectomy patients. The overwhelming majority of bile duct stones are cholesterol stones formed in the gallbladder, which then migrate into the extrahepatic biliary tree through the cystic duct. Primary calculi arising de novo in the ducts are usually brown pigment stones developing in patients with (1) hepatobiliary parasitism or chronic, recurrent cholangitis; (2) congenital anomalies of the bile ducts (especially Caroli's disease); (3) dilated, sclerosed, or strictured ducts; or (4) an *MDR3* (ABCB4) gene defect leading to impaired biliary phospholipids secretion (low phospholipid-associated cholesterol cholelithiasis). Common duct stones may remain asymptomatic for years, may pass spontaneously into the duodenum, or (most often) may present with biliary colic or a complication.

Complications • CHOLANGITIS Cholangitis may be acute or chronic, and symptoms result from inflammation, which usually is caused by at least partial obstruction to the flow of bile. Bacteria are present on bile culture in ~75% of patients with acute cholangitis early in the symptomatic course. The characteristic presentation of acute cholangitis involves biliary pain, jaundice, and spiking fevers with chills (Charcot's triad). Blood cultures are frequently positive, and leukocytosis is typical. *Nonsuppurative acute cholangitis* is most common and may respond relatively rapidly to supportive measures and to treatment with antibiotics. In *suppurative acute cholangitis*, however, the presence of pus under pressure in a completely obstructed ductal system leads to symptoms of severe toxicity—mental confusion, bacteremia, and septic shock. Response to antibiotics alone in this setting is relatively poor, multiple hepatic abscesses are often present, and the mortality rate approaches 100% unless prompt endoscopic or surgical relief of the obstruction and drainage of infected bile are carried out. Endoscopic management of bacterial cholangitis is as effective as surgical intervention. ERCP with endoscopic sphincterotomy is safe and the preferred initial procedure for both establishing a definitive diagnosis and providing effective therapy.

OBSTRUCTIVE JAUNDICE Gradual obstruction of the CBD over a period of weeks or months usually leads to initial manifestations of jaundice or pruritus without associated symptoms of biliary colic or cholangitis. Painless jaundice may occur in patients with choledocholithiasis but is much more characteristic of biliary obstruction secondary to malignancy of the head of the pancreas, bile ducts, or ampulla of Vater.

In patients whose obstruction is secondary to choledocholithiasis, associated chronic calculous cholecystitis is very common, and the gallbladder in this setting may be unable to distend. The absence of a palpable gallbladder in most patients with biliary obstruction from duct stones is the basis for Courvoisier's law, that is, that the presence of a palpably enlarged gallbladder suggests that the biliary obstruction is secondary to an underlying malignancy rather than to calculous disease. Biliary obstruction causes progressive dilatation of the intrahepatic bile ducts as intrabiliary pressures rise. Hepatic bile flow is suppressed, and reabsorption and regurgitation of conjugated bilirubin into the bloodstream lead to jaundice accompanied by dark urine (bilirubinuria) and light-colored (acholic) stools.

CBD stones should be suspected in any patient with cholecystitis whose serum bilirubin level is >85.5 µmol/L (5 mg/dL). The maximum bilirubin level is seldom >256.5 µmol/L (15.0 mg/dL) in patients with choledocholithiasis unless concomitant hepatic or renal disease or another factor leading to marked hyperbilirubinemia exists. Serum bilirubin levels ≥42.0 µmol/L (20 mg/dL) should suggest the possibility of neoplastic obstruction. The serum alkaline phosphatase level is almost always elevated in biliary obstruction. A rise in alkaline phosphatase often precedes clinical jaundice and may be the only abnormality in routine liver function tests. There may be a two- to tenfold elevation of serum aminotransferases, especially in association with acute obstruction. Following relief of the obstructing process, serum aminotransferase elevations usually return rapidly to normal, while the serum bilirubin level may take 1–2 weeks to return to normal.

The alkaline phosphatase level usually falls slowly, lagging behind the decrease in serum bilirubin.

PANCREATITIS The most common associated entity discovered in patients with nonalcoholic acute pancreatitis is biliary tract disease. Biochemical evidence of pancreatic inflammation complicates acute cholecystitis in 15% of cases and choledocholithiasis in >30%, and the common factor appears to be the passage of gallstones through the common duct. Coexisting pancreatitis should be suspected in patients with symptoms of cholecystitis who develop (1) back pain or pain to the left of the abdominal midline, (2) prolonged vomiting with paralytic ileus, or (3) a pleural effusion, especially on the left side. Surgical treatment of gallstone disease is usually associated with resolution of the pancreatitis.

SECONDARY BILIARY CIRRHOSIS Secondary biliary cirrhosis may complicate prolonged or intermittent duct obstruction with or without recurrent cholangitis. Although this complication may be seen in patients with choledocholithiasis, it is more common in cases of prolonged obstruction from stricture or neoplasm. Once established, secondary biliary cirrhosis may be progressive even after correction of the obstructing process, and increasingly severe hepatic cirrhosis may lead to portal hypertension or to hepatic failure and death. Prolonged biliary obstruction may also be associated with clinically relevant deficiencies of the fat-soluble vitamins A, D, E, and K.

Diagnosis and Treatment The diagnosis of choledocholithiasis is made by cholangiography (Table 346-3), either preoperatively by endoscopic retrograde cholangiogram (ERC) (Fig. 346-2C) or MRCP or intraoperatively at the time of cholecystectomy. As many as 15% of patients undergoing cholecystectomy will prove to have CBD stones. When CBD stones are suspected prior to laparoscopic cholecystectomy, preoperative ERCP with endoscopic papillotomy and stone extraction is the preferred approach. It not only provides stone clearance but also defines the anatomy of the biliary tree in relationship to the cystic duct. CBD stones should be suspected in gallstone patients who have any of the following risk factors: (1) a history of jaundice or pancreatitis, (2) abnormal tests of liver function, and (3) ultrasonographic or MRCP evidence of a dilated CBD or stones in the duct. Alternatively, if intraoperative cholangiography reveals retained stones, postoperative ERCP can be carried out. The need for preoperative ERCP is expected to decrease further as laparoscopic techniques for bile duct exploration improve.

The widespread use of laparoscopic cholecystectomy and ERCP has decreased the incidence of complicated biliary tract disease and the need for choledocholithotomy and T-tube drainage of the bile ducts. EBS followed by spontaneous passage or stone extraction is the treatment of choice in the management of patients with common duct stones, especially in elderly or poor-risk patients.

■ TRAUMA, STRICTURES, AND HEMOBILIA

Most benign strictures of the extrahepatic bile ducts result from surgical trauma and occur in about 1 in 500 cholecystectomies. Strictures may present with bile leak or abscess formation in the immediate postoperative period or with biliary obstruction or cholangitis as long as 2 years or more following the inciting trauma. The diagnosis is established by percutaneous or endoscopic cholangiography. Endoscopic brushing of biliary strictures may be helpful in establishing the nature of the lesion and is more accurate than bile cytology alone. When positive exfoliative cytology is obtained, the diagnosis of a neoplastic stricture is established. This procedure is especially important in patients with primary sclerosing cholangitis (PSC) who are predisposed to the development of cholangiocarcinomas. Successful operative correction of non-PSC bile duct strictures by a skillful surgeon with duct-to-bowel anastomosis is usually possible, although mortality rates from surgical complications, recurrent cholangitis, or secondary biliary cirrhosis are high.

Hemobilia may follow traumatic or operative injury to the liver or bile ducts, intraductal rupture of a hepatic abscess or aneurysm of the hepatic artery, biliary or hepatic tumor hemorrhage, or mechanical complications of choledocholithiasis or hepatobiliary parasitism. Diagnostic procedures such as liver biopsy, PTC, and transhepatic biliary

TABLE 346-3 Diagnostic Evaluation of the Bile Ducts

DIAGNOSTIC ADVANTAGES	DIAGNOSTIC LIMITATIONS	CONTRAINdications	COMPLICATIONS	COMMENT
Ultrasound				
Rapid Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts Not limited by jaundice, pregnancy Guidance for fine-needle biopsy	Bowel gas Massive obesity Ascites Barium Partial bile duct obstruction Poor visualization of distal CBD	None	None	Initial procedure of choice in investigating possible biliary tract obstruction
Computed Tomography				
Simultaneous scanning of GB, liver, bile ducts, pancreas Accurate identification of dilated bile ducts, masses Not limited by jaundice, gas, obesity, ascites High-resolution image Guidance for fine-needle biopsy	Extreme cachexia Movement artifact Ileus Partial bile duct obstruction	Pregnancy	Reaction to iodinated contrast, if used	Indicated for evaluation of hepatic or pancreatic masses or for assessing for complications related to gallstones (pancreatitis) Procedure of choice in investigating possible biliary obstruction if diagnostic limitations limit US
Magnetic Resonance Cholangiopancreatography				
Noninvasive modality for visualizing pancreatic and biliary ducts Has excellent sensitivity for bile duct dilatation, biliary stricture, and intraductal abnormalities Can identify pancreatic duct dilatation or stricture, pancreatic duct stenosis, and pancreas divisum	Cannot offer therapeutic intervention High cost	Claustrophobia Certain metals (iron)	None	First choice to assess for choledocholithiasis given comparable sensitivity and specificity to ERCP
Endoscopic Retrograde Cholangiopancreatography				
Simultaneous pancreatography Best visualization of distal biliary tract Bile or pancreatic cytology Endoscopic sphincterotomy and stone removal Biliary manometry	Gastroduodenal obstruction Roux-en-Y biliary-enteric anastomosis	Pregnancy Acute pancreatitis Severe cardiopulmonary disease	Pancreatitis Cholangitis, sepsis Infected pancreatic pseudocyst Perforation (rare) Hypoxemia, aspiration	Cholangiogram of choice if there is believed to be a need for intervention: Diagnosed or high clinical probability of choledocholithiasis Biliary stricture requiring sampling and stenting Need for sphincterotomy such as Sphincter of Oddi dysfunction
Percutaneous Transhepatic Cholangiogram				
Best when bile ducts dilated Best visualization of proximal biliary tract May be used to obtain bile cytology/culture Allows for percutaneous transhepatic drainage	Nondilated or sclerosed ducts	Pregnancy Uncorrectable coagulopathy Massive ascites Hepatic abscess	Bleeding Hemobilia Bile peritonitis Bacteremia, sepsis	Indicated for the drainage of obstructed and infected ducts when ERCP is contraindicated or failed
Endoscopic Ultrasound				
Most sensitive method to detect ampullary stones and exclude pathology in the head of the pancreas				Excellent for detecting choledocholithiasis

Abbreviations: CBD, common bile duct; ERCP, endoscopic retrograde cholangiopancreatography; GB, gallbladder; US, hepatobiliary ultrasound.

drainage catheter placement may also be complicated by hemobilia. Patients often present with a classic triad of biliary pain, obstructive jaundice, and melena or occult blood in the stools. The diagnosis is sometimes made by cholangiographic evidence of blood clot in the biliary tree, but selective angiographic verification may be required. Although minor episodes of hemobilia may resolve without intervention, arteriography and transcatheter embolization or surgical ligation of the bleeding vessel may be required.

■ EXTRINSIC COMPRESSION OF THE BILE DUCTS

Partial or complete biliary obstruction may be produced by extrinsic compression of the ducts. The most common cause of this form of

obstructive jaundice is carcinoma of the head of the pancreas. Biliary obstruction may also occur as a complication of either acute or chronic pancreatitis or involvement of lymph nodes in the porta hepatis by lymphoma or metastatic carcinoma. The latter should be distinguished from cholestasis resulting from massive replacement of the liver by tumor.

■ HEPATOBILIARY PARASITISM

Infestation of the biliary tract by adult helminths or their ova may produce a chronic, recurrent pyogenic cholangitis with or without multiple hepatic abscesses, ductal stones, or biliary obstruction. This condition is relatively rare but does occur in inhabitants of southern China and

2652 elsewhere in Southeast Asia. The organisms most commonly involved are trematodes or flukes, including *Clonorchis sinensis*, *Opisthorchis viverrini* or *Opisthorchis felineus*, and *Fasciola hepatica*. The biliary tract also may be involved by intraductal migration of adult *Ascaris lumbricoides* from the duodenum or by intrabiliary rupture of hydatid cysts of the liver produced by *Echinococcus* spp. The diagnosis is made by cholangiography and the presence of characteristic ova on stool examination. When obstruction is present, the treatment of choice is laparotomy under antibiotic coverage, with common duct exploration and a biliary drainage procedure.

SCLEROSING CHOLANGITIS

PSC is characterized by a progressive, inflammatory, sclerosing, and obliterative process affecting the extrahepatic and/or the intrahepatic bile ducts. PSC is strongly associated with inflammatory bowel disease, especially ulcerative colitis.

Immunoglobulin G4 (IgG4)-associated cholangitis is a biliary disease of unknown etiology that presents with biochemical and cholangiographic features indistinguishable from PSC, is often associated with autoimmune pancreatitis and other fibrosing conditions, and is characterized by elevated serum IgG4 and infiltration of IgG4-positive plasma cells in bile ducts and liver tissue. All patients diagnosed with sclerosing cholangitis should have a serum IgG4 level checked to rule out IgG4 disease as a cause of secondary sclerosing cholangitis, particularly if they do not have inflammatory bowel disease. Glucocorticoids are the initial treatment of choice. Relapse is common after steroid withdrawal, especially with proximal strictures. Long-term treatment with glucocorticoids and/or steroid-sparing agents such as azathioprine may be needed after relapse or for inadequate response (*Chap. 348*).

Patients with PSC often present with signs and symptoms of chronic or intermittent biliary obstruction: RUQ abdominal pain, pruritus, jaundice, or acute cholangitis. Late in the course, complete biliary obstruction, secondary biliary cirrhosis, hepatic failure, or portal hypertension with bleeding varices may occur. The diagnosis is established by finding multifocal, diffusely distributed strictures with intervening segments of normal or dilated ducts, producing a beaded appearance on cholangiography (*Fig. 346-2D*). The cholangiographic techniques of choice in suspected cases are MRCP and ERCP. When a diagnosis of sclerosing cholangitis has been established, causes of secondary sclerosing should be considered. Patients with PSC should undergo testing for associated diseases, especially inflammatory bowel disease, if that diagnosis has not already been established.

Small duct PSC is defined by the presence of chronic cholestasis and hepatic histology consistent with PSC in a patient with IBD, but with normal findings on cholangiography. Small duct PSC is found in ~5% of patients with PSC and may represent an earlier stage of PSC associated with a significantly better long-term prognosis. However, such patients may progress to classic PSC and/or end-stage liver disease with consequent necessity of liver transplantation.

In patients with AIDS, cholangiopancreatography may demonstrate a broad range of biliary tract changes as well as pancreatic duct obstruction and occasionally pancreatitis (*Chap. 202*). Further, biliary tract lesions in AIDS include infection and cholangiopancreatographic changes of sclerosing cholangitis. Changes noted include (1) diffuse involvement of intrahepatic bile ducts alone, (2) involvement of both intra- and extrahepatic bile ducts, (3) ampullary stenosis, (4) stricture of the intrapancreatic portion of the CBD, and (5) pancreatic duct involvement. Associated infectious organisms include *Cryptosporidium*, *Mycobacterium avium-intracellulare*, cytomegalovirus, *Microsporidia*, and *Toxoplasma*. ERCP sphincterotomy can provide significant pain reduction in patients with AIDS-associated papillary stenosis.

TREATMENT

Primary Sclerosing Cholangitis

There is no proven medical therapy for PSC. Therapy to treat pruritus associated with PSC includes cholestyramine, rifampin, and naltrexone. Antibiotics are useful when bacterial cholangitis complicates the clinical picture. Vitamin D and calcium supplementation

may be used as initial therapy to help prevent the loss of bone mass frequently seen in patients with chronic cholestasis. In cases where high-grade biliary obstruction (dominant strictures) has occurred, balloon dilatation is preferred over stenting due to the higher complication rate associated with stenting including pancreatitis and cholangitis. Only rarely is surgical intervention indicated. PSC is a progressive disease with a median survival of 12–18 years following the diagnosis, regardless of therapy. Four variables (age, serum bilirubin level, histologic stage, and splenomegaly) predict survival in patients with PSC and serve as the basis for a risk score. PSC is a common indication for liver transplantation.

FURTHER READING

- B TH et al: Interventional approaches to gallbladder disease. *N Engl J Med* 373:357, 2015.
- L K et al: American College of Gastroenterology (ACG) guidelines: Primary sclerosing cholangitis. *Hepatology* 51:660, 2010.
- R JK et al: Clinical features of acute acalculous cholecystitis. *J Clin Gastroenterol* 36:166, 2003.
- S S: Clinical practice. Acute calculous cholecystitis. *N Engl J Med* 358:2804, 2008.

Section 4 Disorders of the Pancreas

347 Approach to the Patient with Pancreatic Disease

Somashankar G. Krishna,
Darwin L. Conwell, Phil A. Hart



GENERAL CONSIDERATIONS

Globally, pancreatic disorders, including acute and chronic pancreatitis, pancreatic cysts, and pancreatic cancer, are challenging to manage and associated with a high burden on health care resources. The relationships between these diseases continues to be poorly understood, but there is encouraging progress. Acute pancreatitis is one of the most common reasons for hospitalizations in gastroenterology, and there is increasing evidence of long-term sequelae including diabetes, exocrine pancreas insufficiency, and pancreas cancer. Chronic pancreatitis, an irreversible disease of the pancreas, is associated with poor quality of life, largely related to abdominal pain, and associated exocrine insufficiency. Pancreatic cysts, mostly incidental, are increasingly detected on cross-sectional abdominal imaging studies. Although a small number and specific types of pancreatic cysts can progress to pancreatic cancer, diagnostic uncertainty can introduce unwanted anxiety to patients and treating physicians. Meanwhile, with persistently high mortality rates, the incidence of pancreatic adenocarcinoma is increasing and is the seventh leading cause of cancer-related death in the industrialized world and the third most common in the United States.

As emphasized in *Chap. 348*, the etiologies and clinical manifestations of pancreatitis are quite varied. Although it is well-appreciated that acute pancreatitis is frequently secondary to biliary tract disease and alcohol abuse, it can also be caused by drugs, genetic mutations, and trauma. In ~30% of patients with acute pancreatitis and 25–40% of patients with chronic pancreatitis, the etiology is initially unexplained.

The global pooled incidence of acute pancreatitis is ~33.7 cases (95% confidence interval [CI], 23.3–48.8) with 1.16 deaths (95% CI, 0.85–1.6) per 100,000 person-years. The global pooled incidence of chronic pancreatitis is ~9.6 cases (95% CI, 7.9–11.8) with 0.09 attributable deaths (95% CI, 0.02–0.5) per 100,000 person-years. In the

United States, the number of patients admitted to the hospital with acute pancreatitis is increasing, with estimated rates of almost 300,000 annually, whereas the number of patients hospitalized for chronic pancreatitis is decreasing, with recent estimates of ~13,000 admissions per year. Chronic pancreatitis has an annual prevalence of 42–73 cases per 100,000 adults in the United States, although higher prevalence rates (0.04–5%) have been noted among adults at autopsy. Together, acute and chronic pancreatic disease costs an estimated \$3 billion annually in health care expenditures.

The diagnosis of acute pancreatitis is generally clearly defined based on a combination of laboratory, imaging, and clinical symptoms. The diagnosis of chronic pancreatitis, especially in mild disease, is hampered by the relative inaccessibility of the pancreas to direct examination and the nonspecificity of the abdominal pain associated with chronic pancreatitis. Many patients with chronic pancreatitis do not have elevated blood amylase or lipase levels. Some patients with chronic pancreatitis develop signs and symptoms of exocrine pancreatic insufficiency (EPI), and thus, objective evidence for pancreatic disease can be demonstrated. However, there is a very large reservoir of pancreatic exocrine function. More than 90% of the pancreas must be damaged before maldigestion of fat and protein is manifested. Noninvasive, indirect tests of pancreatic exocrine function (e.g., fecal elastase) are much more likely to give abnormal results in patients with obvious advanced pancreatic disease (i.e., pancreatic calcification, steatorrhea, or diabetes mellitus) than in patients with occult disease. Invasive, direct tests of pancreatic secretory function (e.g., secretin stimulation test) are the most sensitive and specific tests to detect early chronic pancreatic disease when imaging is equivocal or normal.

The increasing utilization of cross-sectional imaging modalities with their improved resolution has contributed to a high prevalence (2–5% with computed tomography [CT] scans, 20–30% with magnetic resonance imaging [MRI]) of incidentally detected pancreatic cysts. The most common cyst type encountered is an intraductal papillary mucinous neoplasm (IPMN), which is classified as a precancerous mucinous cyst. In the absence of high-risk features, radiographic surveillance is typically recommended (Fig. 347-1). Mucinous cystic neoplasms (MCNs) are a less commonly encountered mucinous cyst. Among the neoplastic cysts, serous cystadenomas have a negligible risk of progression to malignancy. Other infrequent neoplastic cysts include neuroendocrine tumors and solid pseudopapillary neoplasms. The most commonly encountered benign cyst is a pseudocyst, which can occur in patients with a history of acute or chronic pancreatitis. It is often difficult to accurately predict the risk of malignant transformation of precancerous pancreatic cysts, and there is an increasing number of patients on imaging surveillance protocols burdening the health care systems in the industrialized world.

■ TESTS USEFUL IN THE DIAGNOSIS OF PANCREATIC DISEASE

Several tests are of value in the evaluation of pancreatic disease. Examples of specific tests and their usefulness in the diagnosis of acute and chronic pancreatitis are summarized in Table 347-1 and Fig. 347-2. At some institutions, pancreatic function tests are available and performed if the diagnosis of chronic pancreatitis remains a possibility after noninvasive tests (i.e., ultrasound, CT Scan, MRI with magnetic resonance cholangiopancreatography [MRCP]) or invasive tests (i.e., endoscopic retrograde cholangiopancreatography [ERCP], endoscopic ultrasound [EUS]) have given normal or inconclusive results. In this regard, tests using *direct* stimulation of the pancreas with secretin are the most sensitive.

Pancreatic Enzymes in Body Fluids The serum amylase and lipase levels are widely used as screening tests for acute pancreatitis in the patient with acute abdominal pain or back pain. Lipase is very specific for the pancreas, and values greater than three times the upper limit of normal ($3 \times$ ULN) in combination with epigastric pain strongly suggest the diagnosis of acute pancreatitis. In acute pancreatitis, the serum amylase and lipase are usually elevated within 24 h of onset and remain so for 3–7 days. Levels usually return to normal within 7 days unless there is pancreatic ductal disruption, ductal obstruction, or pseudocyst formation. Approximately 85% of patients with acute pancreatitis have threefold or greater elevated serum lipase and amylase levels. The values may be normal if (1) there is a delay (2–5 days) before blood samples are obtained, (2) the underlying disorder is chronic pancreatitis rather than acute pancreatitis, or (3) hypertriglyceridemia is present. Patients with hypertriglyceridemia and acute pancreatitis have been found to have spuriously low levels of amylase and perhaps lipase activity. In the absence of objective evidence of pancreatitis by abdominal ultrasound, contrast-enhanced CT scan, MRI with MRCP, or EUS, mild to moderate elevations of amylase and/or lipase are not helpful in making a diagnosis of chronic pancreatitis.

It should be noted that the serum amylase can be elevated in other conditions (Table 347-2), in part because the enzyme is found in many organs. In addition to the pancreas and salivary glands, small quantities of amylase are found in the tissues of the fallopian tubes, lung, thyroid, and tonsils and can be produced by various tumors (carcinomas of the lung, esophagus, breast, and ovary). Isoamylase determinations do not accurately distinguish elevated blood amylase levels from pancreatic or nonpancreatic sources. In patients with unexplained hyperamylasemia, the measurement of macroamylase can avoid numerous tests in patients with this rare disorder.

Elevation of ascitic fluid amylase occurs in acute pancreatitis as well as in (1) ascites due to disruption of the main pancreatic duct or a leaking pseudocyst and (2) other abdominal disorders that simulate pancreatitis



FIGURE 347-1 *A.* Side-branch intraductal papillary mucinous neoplasm (magnetic resonance imaging [MRI] with magnetic resonance cholangiopancreatography [MRCP]). T2-weighted MRCP image demonstrates a dominant, lobulated, hyperintense cystic structure (arrow) within the posterior body of the pancreas. The pancreatic duct upstream from the cyst is dilated and irregular. Endoscopic ultrasound and fine-needle aspiration of cyst fluid were consistent with a mucinous cyst. Surgical histopathology revealed an infiltrating moderately differentiated adenocarcinoma, 0.3 cm, arising in a background of an intraductal papillary mucinous neoplasm (IPMN). *B.* Mucinous cystic neoplasm (computed tomography [CT] scan). In the tail of the pancreas, there is a well-circumscribed hypodense cyst (arrow) without any nodular enhancing components. Endoscopic ultrasound and fine-needle aspiration of cyst fluid were suggestive of a mucinous cyst. Surgical histopathology revealed a mucinous cystic neoplasm (3.4 cm) with low-grade dysplasia. The stroma of the cyst demonstrated diffuse positivity for progesterone receptor and focal positivity for CD10 (ovarian stroma), confirming the diagnosis. *C.* Serous cystadenoma (MRI). A lobulated microcystic cyst (arrow) is observed in the tail of the pancreas. Neither a communication with the main pancreatic duct nor intracystic soft tissue enhancing nodular components were observed. However, the cyst continued to increase in size and a distal pancreatectomy was performed. Histopathology revealed a serous microcystic adenoma. (Courtesy of Dr. Z.K. Shah, The Ohio State University Wexner Medical Center; with permission.)

TABLE 347-1 Tests Useful in the Diagnosis of Acute and Chronic Pancreatitis and Pancreatic Neoplasms

TEST	PRINCIPLE	COMMENT
Pancreatic Enzymes in Body Fluids		
Serum lipase	Pancreatic inflammation leads to increased serum enzyme levels	Enzyme measurement of choice for the diagnosis of acute pancreatitis; increased specificity if the level is more than three times the upper limit of normal ($3 \times$ ULN)
Amylase		
1. Serum	Pancreatic inflammation leads to increased serum enzyme levels	Simple; increased specificity if the level is $>3 \times$ ULN; may be falsely normal in patients with hypertriglyceridemic pancreatitis
2. Urine	Renal clearance of amylase is increased in acute pancreatitis	Infrequently used
3. Ascitic fluid	Disruption of gland or main pancreatic duct leads to increased amylase concentration	Can help establish source of ascites; false positives occur with intestinal obstruction and perforated ulcer; can also measure lipase
4. Pleural fluid	Exudative pleural effusion with pancreatitis	False positives occur with carcinoma of the lung and esophageal perforation
Studies Pertaining to Pancreatic Structure		
Radiologic and radionuclide tests		
1. Plain film of the abdomen or upper gastrointestinal x-rays	Can demonstrate large calcifications in chronic pancreatitis	Infrequently used
2. Ultrasonography (US)	Can provide limited information on edema, inflammation, calcification, pseudocysts, and mass lesions	Simple, noninvasive; sequential studies quite feasible; useful in diagnosis of gallstones; pancreas visualization limited by interference from overlying bowel gas
3. Computed tomography (CT) scan	Permits detailed visualization of pancreas and surrounding structures, pancreatic fluid collection, pseudocyst; assessment of necrosis or interstitial disease	Useful in the diagnosis of pancreatic calcification, dilated pancreatic ducts, and pancreatic tumors; may not be able to distinguish between inflammatory and neoplastic mass lesions; multiphasic CT scans are the preferred imaging modality for staging pancreatic cancer; IV contrast is needed for characterization of most features
4. Magnetic resonance imaging (MRI) and cholangiopancreatography (MRCP)	Permits noninvasive detailed evaluation of the pancreatic parenchyma, biliary and pancreatic ducts, adjacent soft tissues, and vascular structures.	Has mostly replaced ERCP for diagnostic assessment of the pancreatic duct; more sensitive than CT scan for detection of mild pancreatitis, necrosis, choledocholithiasis, pancreatic ductal abnormalities, and cystic neoplasms; no exposure to ionizing radiation
5. Endoscopic ultrasonography (EUS) and fine-needle aspiration/biopsy (FNA/B)	High-frequency transducer used with EUS produces very-high-resolution images permitting focused evaluation of pancreatic parenchyma and biliary and pancreatic ducts, and FNA/B provides targeted tissue acquisition	Can be used to assess gallstones, choledocholithiasis, chronic pancreatitis, pancreatic masses, and cystic neoplasms; FNA/B facilitates diagnostic and therapeutic management of pancreatic diseases
6. Endoscopic retrograde cholangiopancreatography (ERCP)	Cannulation of pancreatic and/or common bile duct permits visualization of pancreaticobiliary ductal system	Primarily a therapeutic procedure; invasive with risks for iatrogenic complications
Tests of Exocrine Pancreatic Function		
Direct stimulation of the pancreas with analysis of duodenal contents		
1. Secretin test	Secretin leads to increased output of pancreatic juice and HCO_3^- ; pancreatic secretory response is related to the functional mass of pancreatic tissue; involves duodenal intubation and fluoroscopic placement of gastroduodenal tube	Sensitive to detect occult disease; poorly defined normal enzyme response; large secretory reserve capacity of the pancreas; rarely performed
2. Endoscopic pancreatic function test (ePFT)	Secretin-stimulated collection of pancreatic juice performed during upper endoscopy; replaces need for tube placement in the duodenum	Sensitive to detect occult disease; high negative predictive value for chronic pancreatitis; requires sedation
3. EUS-ePFT	Combines endosonographic evaluation of the pancreas and endoscopic collection of pancreatic juice	Single endoscopic evaluation of pancreatic structure and function
4. Secretin-stimulated MRCP	Combines imaging evaluation of the pancreas and a semiquantitative estimation of pancreatic juice output in the duodenum	Improved visualization of pancreatic ductal anatomy; functional evaluation is less accurate than ePFT; noninvasive
Measurement of intraluminal digestion products		
1. Stool fat determination	Lack of lipolytic enzymes brings about impaired fat digestion; quantitative 72-h stool collection and estimation are more reliable than qualitative analysis of a random stool sample	Reliable reference standard for defining severity of fat malabsorption; does not distinguish between pancreatic and nonpancreatic cause of malabsorption
Measurement of pancreatic enzymes in feces		
1. Fecal elastase	Pancreatic secretion of proteolytic enzymes; not degraded in intestine	Diagnostic accuracy is highest when the pretest probability is high and the value is $<100 \mu\text{g/g}$; false positives will occur in patients with nonformed stools

(e.g., intestinal obstruction, intestinal infarction, or perforated peptic ulcer). Elevation of pleural fluid amylase can occur in acute pancreatitis, chronic pancreatitis, carcinoma of the lung, and esophageal perforation. Lipase is the single best enzyme to measure for the diagnosis of acute pancreatitis. It is important to acknowledge that levels are often mildly

elevated in the setting of renal disease, so determining whether a patient with renal failure and abdominal pain has pancreatitis remains a challenging clinical problem. One study found that serum amylase levels were elevated in patients with renal dysfunction only when creatinine clearance was $<0.8 \text{ mL/s}$ ($<50 \text{ mL/min}$). In such patients, the serum

	<ul style="list-style-type: none"> Clinical signs and symptoms suggestive of chronic pancreatic disease: abdominal pain, nausea, weight loss, steatorrhea, malabsorption, history of alcohol abuse, recurrent pancreatitis, fatty-food intolerance Perform history, physical examination, review of laboratory studies; consider fecal elastase measurement
Step 1	<ul style="list-style-type: none"> Contrast-enhanced CT scan <i>CP diagnostic criteria:</i> calcifications in combination with atrophy and/or dilated duct Diagnostic criteria met; no further imaging needed Inconclusive or nondiagnostic results; continue to step 2
Step 2	<ul style="list-style-type: none"> MRI and MRCP, with or without secretin enhancement (sMRCP) <i>CP diagnostic criteria:</i> Cambridge class III,^a dilated duct, atrophy of gland, filling defects in duct suggestive of stones Diagnostic criteria met; no further imaging needed Inconclusive or nondiagnostic results; continue to step 3
Step 3	<ul style="list-style-type: none"> EUS with quantification of parenchymal and ductal criteria <i>CP diagnostic criteria:</i> ≥5 EUS CP criteria Diagnostic criteria met; no further imaging needed Inconclusive or nondiagnostic results; continue to step 4
Step 4	<ul style="list-style-type: none"> Pancreas function test (with secretin)—endoscopic (ePFT) collection method preferred; consider combining ePFT with EUS <i>CP diagnostic criteria:</i> peak [bicarbonate] <80 mEq/L Diagnostic criteria met; no further imaging needed Inconclusive or nondiagnostic results require monitoring of signs and symptoms and repeat testing in 6 months–1 year

FIGURE 347-2 A stepwise diagnostic approach to the patient with suspected chronic pancreatitis (CP). Endoscopic ultrasonography (EUS) and magnetic resonance imaging (MRI) with secretin-stimulated magnetic resonance cholangiopancreatography (sMRCP/MRCP) are appropriate diagnostic alternatives to endoscopic retrograde cholangiopancreatography (ERCP). CT, computed tomography; ePFT, endoscopic pancreas function test. ^aCambridge classification of pancreatic duct findings: class 0: normal—visualization of complete normal ductal anatomy; class I: equivocal—normal main duct, 1–3 abnormal side branches; class II: mild—normal main duct, >3 abnormal side branches; class III—dilated and irregular main duct, >3 abnormal side branches, small (<10 mm) cysts; class IV—irregular main duct, intraductal calculi, strictures, obstruction with dilation, or large (>10 mm) cysts.

amylase level was invariably <500 IU/L in the absence of objective evidence of acute pancreatitis. In that study, serum lipase and trypsin levels paralleled serum amylase values. With these limitations in mind, the recommended screening test for acute pancreatitis in renal disease is serum lipase, but a high index of clinical suspicion is needed based on symptoms. Elevations in serum lipase >3× ULN due to nonpancreatic etiology can be observed in hepatobiliary or gastrointestinal malignancies, septicemia, liver cirrhosis, systemic lupus erythematosus, severe head injury, chronic alcoholism, diabetes mellitus, and post-ERCP without any associated evidence of pancreatitis.

Studies Pertaining to Pancreatic Structure • RADIOLOGIC TESTS Plain films of the abdomen rarely provide useful information related to pancreatic disease and have been superseded by more detailed imaging studies (ultrasound, EUS, CT, and MRI with MRCP).

Ultrasonography (US) can provide important information in the initial emergency ward evaluation of patients with acute pancreatitis, chronic pancreatitis, pseudocysts, and pancreatic adenocarcinoma. Sonographic changes can indicate the presence of edema, inflammation, and calcification (not obvious on plain films of the abdomen), as well as gallstones, biliary dilation, pseudocysts, and mass lesions. In acute pancreatitis, the pancreas is characteristically enlarged. In pancreatic pseudocyst, the usual appearance is primarily that of a smooth, round fluid collection. Pancreatic adenocarcinoma distorts the usual landmarks, and mass lesions >3.0 cm are usually detected as localized, solid lesions. US is often the initial investigation for most patients with suspected pancreatic disease. However, obesity and excess intestinal bowel gas can interfere with pancreatic imaging, limiting its sensitivity.

CT with intravenous contrast is the best imaging study for the assessment of complications of acute and chronic pancreatitis. It is especially useful in the detection of pancreatic and peripancreatic acute fluid collections, fluid-containing lesions such as pseudocysts, walled-off necrosis (see Chap. 348, Figs. 348-1, 348-2, and 348-4), and pancreatic neoplasms. Acute pancreatitis is characterized by (1) enlargement of the pancreas, (2) distortion of the pancreatic contour with peripancreatic stranding of adjacent fat tissue, and/or (3) the presence of pancreatic fluid that has a different attenuation coefficient than normal pancreas. When possible, CT scans should ideally be performed with oral and intravenous contrast to detect areas of

pancreatic necrosis. The major benefit of CT scan in acute pancreatitis is the diagnosis of pancreatic necrosis in patients not responding to conservative management within 72 h. It may take 48–72 h to develop perfusion defects indicative of pancreatic necrosis. Therefore, if acute pancreatitis is confirmed with serology and physical examination findings, CT scan in the first 3 days is not recommended to minimize risk of contrast-induced nephropathy and unnecessary health care costs. Improved imaging technology and increased resolution are facilitated by multiphasic CT scans using multidetector technology (MDCT) in which a pancreas protocol consisting of dual-phase scanning with intravenous contrast is utilized for the detection and staging of pancreatic cancers. While the sensitivity of MDCT for detecting smaller (≤ 2 cm) lesions is lower, the reported overall sensitivity for pancreatic cancers is 76–97%. The contraindications to using intravenous contrast include renal failure (serum creatinine >2 mg/dL) and a history of severe allergic reaction to iodinated contrast agents. In situations where EUS is not available, CT-guided percutaneous aspiration or biopsy of a pancreatic mass can be performed. Prior to the major advance of EUS-guided fine-needle aspiration (FNA), CT-guided biopsy was utilized in the preceding decades and is regarded as a safe procedure.

MRI and MRCP provide excellent imaging of the bile duct, pancreatic duct, and pancreas parenchyma in both acute pancreatitis and chronic pancreatitis. MRI is better than transabdominal US and CT scans and comparable to EUS in the detection of choledocholithiasis. Similar to CT, MRI can evaluate for the severity of acute pancreatitis. Moreover, T2-weighted MRI of fluid collections can differentiate necrotic debris from fluid in suspected walled-off necrosis, and T1 imaging can diagnose hemorrhage in suspected pseudoaneurysm rupture. In chronic pancreatitis, secretin-enhanced MRCP is a method to enhance the evaluation of major and minor ductal changes. While imaging is comparable to CT for evaluating pancreatic mass lesions, MRI with MRCP is the preferred imaging modality for evaluating pancreatic cystic lesions. Nephrogenic systemic fibrosis has been described in patients with chronic renal failure following exposure to the gadolinium contrast, but incidence rates are extraordinarily low with contemporary contrast agents.

EUS produces high-resolution images of the bile duct, pancreatic parenchyma, and pancreatic duct with a transducer fixed to an

TABLE 347-2 Causes of Hyperamylasemia and Hyperamylasuria

Pancreatic Disease

- I. Pancreatitis
 - A. Acute
 - B. Chronic: ductal obstruction
 - C. Complications of pancreatitis
 1. Pancreatic pseudocyst
 2. Ascites caused by pancreatic duct disruption
 3. Pancreatic necrosis
- II. Pancreatic trauma
- III. Pancreatic adenocarcinoma

Nonpancreatic Disorders

- I. Renal insufficiency
- II. Salivary gland lesions
 - A. Mumps
 - B. Calculus
 - C. Irradiation sialadenitis
 - D. Maxillofacial surgery
- III. "Tumor" hyperamylasemia
 - A. Carcinoma of the lung, esophagus, breast, or ovary
- IV. Macroamylasemia
- V. Burns
- VI. Diabetes mellitus, particularly when ketoacidosis is present
- VII. Pregnancy
- VIII. Renal transplantation
- IX. Cerebral trauma
- X. Drugs: opiates

Other Abdominal Disorders

- I. Biliary tract disease: cholecystitis, choledocholithiasis
- II. Intraabdominal disease
 - A. Perforated or penetrating peptic ulcer
 - B. Intestinal obstruction or inflammation
 - C. Ruptured ectopic pregnancy
 - D. Peritonitis
 - E. Aortic aneurysm
 - F. Postoperative hyperamylasemia

endoscope that can be directed onto the surface of the pancreas through the stomach or duodenum. EUS is not beneficial for the evaluation of pancreas during acute pancreatitis. It is preferable to perform EUS after the resolution of acute pancreatitis (~4 weeks) to detect any predisposing factors, including malignancy, choledocholithiasis, pancreatic divisum, or ampullary lesions. EUS can be combined with ERCP in a single session and is increasingly preferred for the diagnosis and management of choledocholithiasis in acute pancreatitis and pancreatic neoplasm with biliary obstruction. EUS has been studied as a diagnostic modality for chronic pancreatitis. Criteria for abnormalities on EUS in severe chronic pancreatic disease have been developed. There is general agreement that the presence of five or more of the nine criteria listed in Table 347-3 is highly predictive of chronic pancreatitis in the correct clinical context. The sensitivity of EUS (81%; 95% CI, 70–89%) to diagnose chronic pancreatitis is comparable to that of MRI/MRCP (78%; 95% CI, 69–85%) and better than CT (75%; 95% CI, 66–83%); however, nonspecific changes are commonly seen in the pancreas that may be attributable to cigarette smoking, diabetes, or normal aging. EUS also facilitates the delivery of nerve-blocking agents via fine-needle injection in patients suffering from pancreatic pain from chronic pancreatitis (celiac plexus block) or cancer (celiac plexus neurolysis). When clinically suspected, EUS imaging is more sensitive than MDCT for the detection of pancreatic malignancy and permits fine-needle aspiration/biopsy (FNA/B). Currently, EUS-guided FNA/B is the diagnostic modality of choice for the acquisition of diagnostic tissue and cyst fluid in patients with pancreatic masses and cystic lesions, respectively.

TABLE 347-3 Endoscopic Ultrasonographic Criteria for Chronic Pancreatitis (Total Criteria = 9)

DUCTAL	PARENCHYMAL
Stones	Echogenic strands
Hyperechoic main duct margins	Echogenic foci
Main duct irregularity	Lobular contour
Main duct dilatation	Cysts
Visible side branches	

Although a pancreateogram during ERCP is the most specific and sensitive test for evaluating the ductal anatomy, EUS and MRI/MRCP have largely replaced ERCP in the diagnostic evaluation of pancreatic disease to avoid the risk of complications. Therefore, ERCP is primarily of therapeutic value after CT, EUS, or MRI and MRCP has detected abnormalities requiring endoscopic treatment. ERCP is the most sensitive modality for the detection of bile duct stones. In the management of acute biliary pancreatitis, ERCP should not be unduly delayed in patients with high clinical suspicion of biliary obstruction. In chronic pancreatitis, ERCP abnormalities in the main pancreatic duct and side branches have been outlined by the Cambridge classification (Fig. 347-2). The presence of ductal stenosis and irregularity can make it difficult to distinguish chronic pancreatitis from pancreatic adenocarcinoma. It is important to be aware that ERCP changes interpreted as indicating chronic pancreatitis actually may be due to the effects of aging on the pancreatic duct, sequelae of a recent attack of acute pancreatitis, or changes secondary to placement of pancreatic duct stent. Although aging may cause impressive ductal alterations, it does not affect the results of pancreatic secretin function tests. Pancreatic adenocarcinoma is characterized by stenosis or obstruction of either the pancreatic duct or the common bile duct; both ductal systems are often abnormal (double-duct sign). When indicated, ERCP permits acquisition of diagnostic tissue as in biopsy of ampullary lesions or biliary brushings for distal bile duct strictures. Elevated serum amylase levels after ERCP have been reported in the majority of patients, and clinical pancreatitis has been reported in 5–10% of patients. Until recently, pancreatic duct stents were commonly placed to prevent post-ERCP pancreatitis. However, recent data suggest that periprocedural administration of rectal indomethacin can decrease the incidence of post-ERCP pancreatitis. Studies are currently underway comparing rectal indomethacin alone versus combination with prophylactic pancreatic duct stents to prevent post-ERCP pancreatitis.

■ TESTS OF EXOCRINE PANCREATIC FUNCTION

Pancreatic function tests (Table 347-1) can be divided into the following:

1. *Direct stimulation of the pancreas* by IV infusion of secretin followed by collection and measurement of duodenal contents: The secretin test, used to detect diffuse pancreatic disease, is based on the physiologic principle that the pancreatic secretory response is directly related to the functional mass of pancreatic tissue. In the standard assay, secretin is given IV in a dose of 0.2 µg/kg of synthetic human secretin as a bolus. Normal values for the standard secretin test are (1) volume output >2 mL/kg per h, (2) bicarbonate (HCO_3^-) concentration >80 mmol/L, and (3) HCO_3^- output >10 mmol/L in 1 h. The most reproducible measurement, giving the highest level of discrimination between normal subjects and patients with chronic pancreas dysfunction, appears to be the maximal bicarbonate concentration. A cutoff point <80 mmol/L is considered abnormal and suggestive of reduced secretory function that is most commonly observed in early chronic pancreatitis.
2. There may be a dissociation between the results of the secretin test and other tests of absorptive function. For example, patients with chronic pancreatitis often have abnormally low outputs of HCO_3^- after secretin but have normal fecal fat excretion. The secretin test directly measures the secretory capacity of ductular epithelium, whereas fecal fat excretion indirectly reflects intraluminal lipolytic

activity. Steatorrhea does not occur until intraluminal levels of lipase are markedly reduced, underscoring the fact that only small amounts of enzymes are necessary for intraluminal digestive activities. It must be emphasized that an abnormal secretin test result suggests that pancreatic ductal secretory function is abnormal. This is an early abnormality in chronic pancreatitis but should not be considered diagnostic and must be interpreted within the proper clinical context (for example, a patient with recurrent attacks of pancreatitis and persistent abdominal pain and Cambridge 2 changes on imaging).

3. *Measurement of fecal pancreatic enzymes* such as elastase: Measurement of *intraluminal digestion products* (i.e., undigested muscle fibers, stool fat, and fecal nitrogen) is discussed in Chap. 325. The amount of human elastase in stool reflects the pancreatic output of this proteolytic enzyme. Decreased fecal elastase-1 (FE-1) activity in stool is a test to detect severe EPI in patients with chronic pancreatitis and cystic fibrosis. FE-1 levels >200 µg/g are normal, levels of 100–200 µg/g are considered mild-moderate EPI, and levels <100 µg/g are severe EPI. Although the test is simple and noninvasive, it can yield false-positive results if stools are not formed and should not generally be used for the evaluation of a patient with diarrhea. False-positive results have also been observed in diabetes and irritable bowel syndrome.

Tests useful in the diagnosis of EPI and the differential diagnosis of malabsorption are also discussed in Chaps. 325 and 348.

FURTHER READING

- C DL et al: American Pancreatic Association practice guidelines in chronic pancreatitis: Evidence-based report on diagnostic guidelines. *Pancreas* 43:1143, 2014.
 H PA et al: Endoscopic pancreas fluid collection: Methods and relevance for clinical care and translational science. *Am J Gastroenterol* 111:1258, 2016.
 P MS, Y D: Global epidemiology and holistic prevention of pancreatitis. *Nat Rev Gastroenterol Hepatol* 16:175, 2019.
 S VK et al: Diagnosis and management of chronic pancreatitis: A review. *JAMA* 322:2422, 2019.

(tryptophan, phenylalanine, valine, methionine), and gastric acid itself. CCK evokes an enzyme-rich secretion from acinar cells in the pancreas. The *parasympathetic nervous system* (via the vagus nerve) exerts significant control over pancreatic secretion, particularly during the cephalic phase. Secretion evoked by secretin and CCK depends on the permissive roles of vagal afferent and efferent pathways. This is particularly true for enzyme secretion, whereas water and bicarbonate secretions are heavily dependent on the hormonal effects of secretin and to a lesser extent CCK. Also, vagal stimulation affects the release of vasoactive intestinal peptide (VIP), a secretin agonist. Pancreatic exocrine secretion is also influenced by inhibitory neuropeptides including somatostatin, pancreatic polypeptide, peptide YY, neuropeptide Y, enkephalin, pancreastatin, calcitonin gene-related peptides, glucagon, and galanin. Pancreatic polypeptide and peptide YY may act primarily on nerves outside the pancreas, while somatostatin acts at multiple sites.

WATER AND ELECTROLYTE SECRETION

Bicarbonate is the ion of primary physiologic importance within pancreatic secretion. The ductal cells secrete bicarbonate predominantly derived from plasma (93%) more than from intracellular metabolism (7%). Bicarbonate enters the duct lumen through the sodium bicarbonate cotransporter with depolarization caused by chloride efflux through the cystic fibrosis transmembrane conductance regulator (CFTR). Secretin and VIP bind at the basolateral surface and cause an increase in secondary messenger intracellular cyclic AMP and act on the apical surface of the ductal cells opening the CFTR, which promotes secretion. CCK, acting as a neuromodulator, markedly potentiates the stimulatory effects of secretin. Acetylcholine also plays an important role in ductal cell secretion. Intraluminal bicarbonate secreted from the ductal cells helps neutralize gastric acid, increases the solubility of fatty acids and bile acids, maintains an optimal pH for pancreatic and brush border enzymes, and prevents intestinal mucosal damage.

ENZYME SECRETION

The acinar cell is highly compartmentalized for the production and secretion of pancreatic enzymes. Proteins synthesized by the rough endoplasmic reticulum are processed in the Golgi and then targeted to the appropriate site: zymogen granules, lysosomes, or other cell compartments. The zymogen granules migrate to the apical region of the acinar cell awaiting the appropriate neural or hormonal stimulatory response. The pancreas secretes amylolytic, lipolytic, and proteolytic enzymes into the duct lumen. *Amylolytic enzymes*, such as amylase, hydrolyze starch to oligosaccharides and to the disaccharide maltose. The *lipolytic enzymes* include lipase, phospholipase A₁, and cholesterol esterase. Bile salts inhibit lipase in isolation, but colipase, another constituent of pancreatic secretion, binds to lipase and prevents this inhibition. Bile salts activate phospholipase A and cholesterol esterase. *Proteolytic enzymes* include endopeptidases (trypsin, chymotrypsin), which act on internal peptide bonds of proteins and polypeptides; exopeptidases (carboxypeptidases, aminopeptidases), which act on the free carboxyl- and amino-terminal ends of peptides, respectively; and elastase. The proteolytic enzymes are secreted as inactive zymogen precursors. Ribonucleases (deoxyribonucleases, ribonuclease) are also secreted. *Enterokinase*, an enzyme found in the duodenal mucosa ("brush border"), cleaves the lysine-isoleucine bond of trypsinogen to form trypsin. Trypsin then activates the other proteolytic zymogens and phospholipase A₂ in a cascade. The nervous system initiates pancreatic enzyme secretion. The neurologic stimulation is cholinergic, involving extrinsic innervation by the vagus nerve and subsequent innervation by intrapancreatic cholinergic nerves. The stimulatory neurotransmitters are acetylcholine and gastrin-releasing peptides. These neurotransmitters activate calcium-dependent secondary messenger systems, resulting in the release of zymogens into the pancreas duct. VIP is present in intrapancreatic nerves and potentiates the effect of acetylcholine. In contrast to other species, there are no CCK receptors on acinar cells in humans. CCK in physiologic concentrations stimulates pancreatic secretion by stimulating afferent vagal and intrapancreatic nerves.

348

Acute and Chronic Pancreatitis

Phil A. Hart, Darwin L. Conwell,
Somashekhar G. Krishna

BIOCHEMISTRY AND PHYSIOLOGY OF PANCREATIC EXOCRINE SECRETION

GENERAL CONSIDERATIONS

The pancreas secretes 1500–3000 mL of isosmotic alkaline (pH >8) fluid per day containing ~20 enzymes. Pancreatic secretions provide the enzymes and bicarbonate needed to perform the major digestive activity of the gastrointestinal tract and provide an optimal pH for the function of these enzymes.

REGULATION OF PANCREATIC SECRETION

Secretions from the exocrine pancreas are highly regulated by neurohormonal systems in a phasic manner (cephalic, gastric, and intestinal phases). *Gastric acid* is the stimulus for the release of secretin from the duodenal mucosa (S cells), which stimulates the secretion of water and electrolytes from pancreatic ductal cells. Release of cholecystokinin (CCK) from the duodenal and proximal jejunal mucosa (Ito cells) is largely triggered by long-chain fatty acids, essential amino acids

Autodigestion of the pancreas is prevented by (1) the packaging of pancreatic proteases in the precursor (proenzyme) form, (2) intracellular calcium homeostasis (low intracellular calcium in the cytosol of the acinar cell promotes the destruction of spontaneously activated trypsin), (3) acid-base balance, and (4) the synthesis of protective protease inhibitors (pancreatic secretory trypsin inhibitor [PSTI] or SPINK1), which can bind and inactivate ~20% of intracellular trypsin activity. Chymotrypsin C can also lyse and inactivate trypsin. These protease inhibitors are found in acinar cells, pancreatic secretions, and the α_1 - and α_2 -globulin fractions of plasma. Loss of any of these four protective mechanisms leads to premature enzyme activation, autodigestion, and ultimately acute pancreatitis.

■ ENTEROPANCREATIC AXIS AND FEEDBACK INHIBITION

Pancreatic enzyme secretion is controlled, at least in part, by a negative feedback mechanism induced by the presence of active serine proteases in the duodenum and nutrients in the distal small intestine. For example, perfusion of the duodenal lumen with phenylalanine (stimulates early digestion) causes a prompt increase in plasma CCK levels as well as increased secretion of chymotrypsin and other pancreatic enzymes. However, simultaneous perfusion with trypsin (stimulates late digestion) blunts both responses. Conversely, perfusion of the duodenal lumen with protease inhibitors actually leads to enzyme hypersecretion. Available evidence supports the concept that the duodenum contains a peptide called *CCK-releasing factor* (CCK-RF) that is involved in stimulating CCK release. It appears that serine proteases inhibit pancreatic secretion by inactivating a CCK-releasing peptide in the lumen of the small intestine. Thus, the integrative result of both bicarbonate and enzyme secretion depends on a feedback process for both bicarbonate and pancreatic enzymes. Acidification of the duodenum releases secretin, which stimulates vagal and other neural pathways to activate pancreatic duct cells, which secrete bicarbonate. This bicarbonate then neutralizes the duodenal acid, and the feedback loop is completed. Dietary proteins bind proteases, thereby leading to an increase in free CCK-RF. CCK is then released into the blood in physiologic concentrations, acting primarily through the neural pathways (vagal-vagal). This leads to acetylcholine-mediated pancreatic enzyme secretion. Proteases continue to be secreted from the pancreas until the protein within the duodenum is digested. At this point, pancreatic protease secretion is reduced to basic levels, thus completing this step in the feedback process. Additional hormonal feedback inhibition of pancreatic enzyme secretion occurs via peptide YY and glucagon-like peptide-1 following lipid or carbohydrate exposure to the ileum.

ACUTE PANCREATITIS

■ GENERAL CONSIDERATIONS

Recent U.S. estimates indicate that acute pancreatitis is the most common inpatient principal gastrointestinal diagnosis, responsible for >250,000 hospitalizations per year. The annual incidence ranges from 15–45/100,000 persons, depending on the distribution of etiologies (e.g., alcohol, gallstones, metabolic factors, drugs [Table 348-1]) and country of study. The median length of hospital stay is 4 days, with a median hospital cost of ~\$6000 and a mortality of ~1%. The estimated cost annually approaches \$3 billion. Hospitalization rates increase with age and are higher among blacks and men. The age-adjusted rate of hospital discharges with an acute pancreatitis diagnosis increased by 62% between 1988 and 2004. From 2000 to 2009, the rate increased by 30%. Thus, the incidence of acute pancreatitis continues to rise and is associated with substantial health care costs.

■ ETIOLOGY AND PATHOGENESIS

There are many causes of acute pancreatitis (Table 348-1), and the mechanisms by which each of these conditions triggers pancreatic inflammation have not been fully elucidated. Gallstones and alcohol account for 80–90% of identified cases of acute pancreatitis in the United States. Gallstones continue to be the leading cause of acute

TABLE 348-1 Causes of Acute Pancreatitis

Common Causes	
Gallstones (including microlithiasis)	
Alcohol (acute and chronic alcoholism)	
Hypertriglyceridemia	
Endoscopic retrograde cholangiopancreatography (ERCP), especially after biliary manometry	
Idiopathic	
Uncommon Causes	
Drugs (azathioprine, 6-mercaptopurine, sulfonamides, estrogens, tetracycline, valproic acid, 5-aminosalicylic acid [5-ASA])	
Connective tissue disorders and thrombotic thrombocytopenic purpura (TTP)	
Pancreatic cancer	
Hypercalcemia	
Periampullary diverticulum	
Pancreas divisum ^a	
Hereditary pancreatitis	
Cystic fibrosis	
Renal failure	
Infections (mumps, coxsackievirus, cytomegalovirus, echovirus, parasites)	
Autoimmune (e.g., type 1 and type 2)	
Trauma (especially blunt abdominal trauma)	
Postoperative (abdominal and nonabdominal operations)	
Causes to Consider in Patients with Recurrent Bouts of Acute Pancreatitis without an Obvious Etiology	
Occult disease of the biliary tree or pancreatic ducts, especially microlithiasis, biliary sludge	
Alcohol abuse	
Metabolic: Hypertriglyceridemia, hypercalcemia	
Anatomic: Pancreas divisum ^a	
Pancreatic cancer	
Intraductal papillary mucinous neoplasm (IPMN)	
Hereditary pancreatitis	
Cystic fibrosis	
Idiopathic	

^aPancreas divisum is not believed to cause acute pancreatitis in isolation of another disease precipitant.

pancreatitis in most series (30–60%). The risk of acute pancreatitis in patients with at least one gallstone <5 mm in diameter is four-fold greater than that in patients with larger stones. Alcohol is the second most common cause, responsible for 15–30% of cases in the United States. The incidence of pancreatitis in alcoholics is surprisingly low (5/100,000), indicating that in addition to the amount of alcohol ingested, other factors affect a person's susceptibility to pancreatic injury, such as cigarette smoking and genetic predisposition. Acute pancreatitis occurs in 5–10% of patients following endoscopic retrograde cholangiopancreatography (ERCP); however, this risk can be decreased with proper patient selection and the use of a prophylactic pancreatic duct stent and/or rectal nonsteroidal anti-inflammatory drugs (NSAIDs; indomethacin). Risk factors for post-ERCP pancreatitis include minor papilla sphincterotomy, suspected sphincter of Oddi dysfunction, prior history of post-ERCP pancreatitis, age <60 years, more than two contrast injections into the pancreatic duct, and endoscopist experience.

Hypertriglyceridemia is the cause of acute pancreatitis in 1–4% of cases; serum triglyceride levels are usually >1000 mg/dL. Most patients with hypertriglyceridemic pancreatitis have undiagnosed or uncontrolled diabetes mellitus. An additional subset has an underlying derangement in lipid metabolism, probably unrelated to pancreatitis. Such patients are prone to recurrent episodes of pancreatitis. Any factor (e.g., alcohol or medications, such as oral contraceptives) that causes an abrupt increase in serum triglycerides can potentially precipitate a bout

of acute pancreatitis. Patients with a deficiency of apolipoprotein CII have an increased incidence of pancreatitis; apolipoprotein CII activates lipoprotein lipase, which is important in clearing chylomicrons from the bloodstream. Although frequently entertained, <2% of cases of acute pancreatitis are drug related. Drugs cause pancreatitis either by a hypersensitivity reaction or by the generation of a toxic metabolite, although in some cases, it is not clear which of these mechanisms is operative (Table 348-1).

Pathologically, acute pancreatitis ranges from *interstitial pancreatitis* (pancreas blood supply maintained), which is generally self-limited, to *necrotizing pancreatitis* (pancreas blood supply interrupted). Autodigestion is a currently accepted pathogenic theory resulting when proteolytic enzymes (e.g., trypsinogen, chymotrypsinogen, proelastase, and lipolytic enzymes such as phospholipase A₂) are activated in the pancreas acinar cell compartment rather than the intestinal lumen. A number of factors (e.g., endotoxins, exotoxins, viral infections, ischemia, oxidative stress, lysosomal calcium, direct trauma) are believed to facilitate premature activation of trypsin. Activated proteolytic enzymes, especially trypsin, not only digest pancreatic and peripancreatic tissues but can also activate other enzymes, such as elastase and phospholipase A₂. Spontaneous activation of trypsin also can occur, resulting in autodigestion.

■ ACTIVATION OF PANCREATIC ENZYMES IN THE PATHOGENESIS OF ACUTE PANCREATITIS

Several studies have suggested that pancreatitis is a disease that evolves in three phases. The *initial phase* is characterized by intrapancreatic digestive enzyme activation and acinar cell injury. Trypsin activation appears to be mediated by lysosomal hydrolases such as cathepsin B that become colocalized with digestive enzymes in intracellular organelles; it is currently believed that acinar cell injury is the consequence of trypsin activation. The *second phase* of pancreatitis involves the activation, chemoattraction, and sequestration of leukocytes and macrophages in the pancreas, resulting in an enhanced intrapancreatic inflammatory reaction. Neutrophil depletion induced by prior administration of an antineutrophil serum has been shown to reduce the severity of experimentally induced pancreatitis. There is also evidence to support the concept that neutrophils can activate trypsinogen. Thus, intrapancreatic acinar cell activation of trypsinogen could be a two-step process (i.e., an early neutrophil-independent and a later neutrophil-dependent phase). The *third phase* of pancreatitis is due to the effects of activated proteolytic enzymes and cytokines, released by the inflamed pancreas, on distant organs. Activated proteolytic enzymes, especially trypsin, not only digest pancreatic and peripancreatic tissues but also activate other enzymes such as elastase and phospholipase A₂. The active enzymes and cytokines then digest cellular membranes and cause proteolysis, edema, interstitial hemorrhage, vascular damage, coagulation necrosis, fat necrosis, and cellular necrosis in the parenchyma. Cellular injury and death result in the liberation of bradykinin peptides, vasoactive substances, and histamine that can produce vasodilation, increased vascular permeability, and edema with profound effects on other organs. The systemic inflammatory response syndrome (SIRS) and acute respiratory distress syndrome (ARDS), as well as multiorgan failure, may occur as a result of this cascade of local and distant effects.

A number of genetic factors can increase the susceptibility and/or modify the severity of pancreatic injury in acute pancreatitis, recurrent acute pancreatitis, and chronic pancreatitis. All the major genetic susceptibility factors center on the control of trypsin activity within the pancreatic acinar cell, in part because they were identified as candidate genes linked to intrapancreatic trypsin control. Six genetic variants have been identified as being associated with susceptibility to pancreatitis. The genes that have been identified include (1) cationic trypsinogen gene (*PRSS1*), (2) pancreatic secretory trypsin inhibitor (*SPINK1*), (3) the cystic fibrosis transmembrane conductance regulator gene (*CFTR*), (4) the chymotrypsin C gene (*CTRC*), (5) the calcium-sensing receptor (*CASR*), and (6) claudin-2 (*CLDN2*). Among these variants, only *PRSS1* mutations are sufficient to precipitate acute pancreatitis in the absence of other risk factors, whereas the other variants are disease

modifiers. Investigations of other genetic variants are currently underway, and new genes will be added to this list in the future.

APPROACH TO THE PATIENT

Abdominal Pain

Abdominal pain is the major symptom of acute pancreatitis. Pain may vary from mild discomfort to severe, constant, and incapacitating distress. Characteristically, the pain, which is steady and boring in character, is located in the epigastrium region and may radiate to the back, chest, flanks, and lower abdomen. Nausea, vomiting, and abdominal distension due to gastric and intestinal hypomotility are also frequent complaints.

Physical examination frequently reveals a distressed and anxious patient. Low-grade fever, tachycardia, and hypotension are common. Shock is not unusual and may result from (1) hypovolemia secondary to exudation of blood and plasma proteins into the retroperitoneal space; (2) increased formation and release of kinin peptides, which cause vasodilation and increased vascular permeability; and (3) systemic effects of proteolytic and lipolytic enzymes released into the circulation. Jaundice occurs infrequently; when present, it may be a consequence of extrinsic compression due to peripancreatic edema or a pancreatic head mass or of intraductal obstruction from a common bile duct stone or sludge. Erythematous skin nodules due to subcutaneous fat necrosis rarely occur. In 10–20% of patients, there are pulmonary findings, including basilar rales, atelectasis, and pleural effusion, the latter most frequently left-sided. Abdominal tenderness and muscle rigidity are present to a variable degree, but compared with the intense pain, these signs may be less impressive. Bowel sounds are usually diminished or absent. An enlarged pancreas from an acute fluid collection, walled-off necrosis, or a pseudocyst may be palpable in the upper abdomen later in the course of the disease (i.e., 4–6 weeks). A faint blue discoloration around the umbilicus (Cullen's sign) may occur as the result of hemoperitoneum, and a blue-red-purple or green-brown discoloration of the flanks (Turner's sign) reflects tissue breakdown of hemoglobin from severe necrotizing pancreatitis with hemorrhage; both findings are rare but reflect an increased clinical severity.

■ LABORATORY DATA

Serum amylase and lipase values threefold or more above normal are strongly supportive of the diagnosis if alternate etiologies, including gut perforation, ischemia, and infarction, are excluded. However, it should be noted that there is no correlation between the severity of pancreatitis and the degree of serum lipase and amylase elevations or serial trends. After 3–7 days, even with continuing evidence of pancreatitis, total serum amylase values tend to return toward normal. However, pancreatic lipase levels may remain elevated for 7–14 days. It should be recognized that amylase elevations in serum and urine occur in many conditions other than pancreatitis (see Chap. 347, Table 347-2). Importantly, patients with *acidemia* (arterial pH ≤ 7.32) may have spurious elevations in serum amylase. This finding explains why patients with diabetic ketoacidosis may have marked elevations in serum amylase without any other evidence of acute pancreatitis. On the other hand, serum amylase levels can be spuriously low in the setting of severe hypertriglyceridemia. Serum lipase activity increases in parallel with amylase activity and is more specific than amylase, making it the preferred test. A serum lipase measurement can be instrumental in differentiating a pancreatic or nonpancreatic cause for hyperamylasemia.

Leukocytosis (15,000–20,000 leukocytes/ μL) occurs frequently. Patients with more severe disease may show hemoconcentration with hematocrit values $>44\%$ and/or prerenal azotemia with a blood urea nitrogen (BUN) level $>22 \text{ mg/dL}$ resulting from loss of plasma into the retroperitoneal space and peritoneal cavity.

Hemoconcentration may be the harbinger of more severe disease, whereas azotemia is a significant risk factor for mortality. **Hyperglycemia** is common and is due to multiple factors, including

TABLE 348-2 Revised Atlanta Definitions of Morphologic Features of Acute Pancreatitis

	DEFINITION	COMPUTED TOMOGRAPHY FEATURES
Types of Acute Pancreatitis		
Interstitial pancreatitis	Acute inflammation of the pancreatic parenchyma and peripancreatic tissues, but without recognizable tissue necrosis	Pancreatic parenchyma enhancement by IV contrast agent and without peripancreatic necrosis
Necrotizing pancreatitis	Inflammation associated with pancreatic parenchymal and/or peripancreatic necrosis	Lack of pancreatic parenchymal enhancement by IV contrast agent and/or presence of findings of peripancreatic necrosis (see below—ANC and WON)
Morphologic Features		
Acute pancreatic fluid collection	Peripancreatic fluid associated with interstitial edematous pancreatitis with no associated peripancreatic necrosis. This term applies only to areas of peripancreatic fluid seen within the first 4 weeks after onset of interstitial edematous pancreatitis and without the features of a pseudocyst.	Occurs in the setting of interstitial pancreatitis Homogeneous collection with fluid density Confined by normal peripancreatic fascial planes No definable wall encapsulating the collection Adjacent to pancreas (no intrapancreatic extension)
Pancreatic pseudocyst	An encapsulated collection of fluid with a well-defined inflammatory wall usually outside the pancreas with minimal or no necrosis. This entity usually occurs >4 weeks after onset of interstitial edematous pancreatitis.	Well circumscribed, usually round or oval Homogeneous fluid density No solid component Well-defined wall; that is, completely encapsulated Maturation usually requires >4 weeks after onset of acute pancreatitis; occurs after interstitial pancreatitis
Acute necrotic collection (ANC)	A collection containing variable amounts of both fluid and necrosis associated with necrotizing pancreatitis; the necrosis can involve the pancreatic parenchyma and/or the peripancreatic tissues.	Occurs in the setting of acute necrotizing pancreatitis Heterogeneous and nonliquid density of varying degrees in different locations (some appear homogeneous early in their course) No definable wall encapsulating the collection Location—intrapancreatic and/or extrapancreatic
Walled-off necrosis (WON)	A mature, encapsulated collection of pancreatic and/or peripancreatic necrosis that has developed a well-defined inflammatory wall. WON usually occurs >4 weeks after onset of acute necrotizing pancreatitis.	Heterogeneous with liquid and nonliquid density with varying degrees of loculations (some may appear homogeneous) Well-defined wall; that is, completely encapsulated Location—intrapancreatic and/or extrapancreatic Maturation usually requires >4 weeks after onset of acute necrotizing pancreatitis

Source: Data from P Banks et al: Gut 62:102, 2013.

decreased insulin release, increased glucagon release, and increased output of adrenal glucocorticoids and catecholamines. *Hypocalcemia* occurs in ~25% of patients, and its pathogenesis is incompletely understood. Although earlier studies suggested that the response of the parathyroid gland to a decrease in serum calcium is impaired, subsequent observations have failed to confirm this phenomenon. Intraperitoneal saponification of calcium by fatty acids in areas of fat necrosis occurs occasionally, with large amounts (up to 6.0 g) dissolved or suspended in ascitic fluid. Such “soap formation” may also be significant in patients with pancreatitis, mild hypocalcemia, and little or no obvious ascites. *Hyperbilirubinemia* (serum bilirubin >4.0 mg/dL) occurs in ~10% of patients. However, jaundice is transient, and serum bilirubin levels return to normal in 4–7 days. Serum alkaline phosphatase and transaminase levels may also be transiently elevated and parallel serum bilirubin values. Elevations of alanine aminotransferase (ALT) >3x the upper limit of normal are strongly associated with a gallstone etiology in patients with acute pancreatitis. Approximately 5–10% of patients have *hypoxemia* (arterial P_2 <60 mmHg), which may herald the onset of ARDS. Finally, the electrocardiogram is occasionally abnormal in acute pancreatitis with ST-segment and T-wave abnormalities simulating myocardial ischemia.

An abdominal ultrasound is recommended in the emergency ward as the initial diagnostic imaging modality and is most useful to evaluate for gallstones and common bile duct dilation.

The Revised Atlanta Criteria have clearly outlined the morphologic features of acute pancreatitis on computed tomography (CT) scan as follows: (1) interstitial pancreatitis, (2) necrotizing pancreatitis, (3) acute pancreatic fluid collection, (4) pancreatic pseudocyst, (5) acute necrotic collection (ANC), and (6) walled-off necrosis (WON) (Table 348-2 and Fig. 348-1). Radiologic studies useful in the diagnosis of acute pancreatitis are discussed in Chap. 347 and listed in Table 347-1.

■ DIAGNOSIS

Any severe acute pain in the abdomen or back should suggest the possibility of acute pancreatitis. The diagnosis is established by two of the following three criteria: (1) typical abdominal pain in the epigastrium that may radiate to the back, (2) threefold or greater elevation in serum lipase and/or amylase, and (3) confirmatory findings of acute pancreatitis on cross-sectional abdominal imaging. Although not required for diagnosis, markers of severity may include hemoconcentration (hematocrit >44%), admission azotemia (BUN >22 mg/dL), SIRS, and signs of organ failure (Table 348-3).

The *differential diagnosis* should include the following disorders: (1) perforated viscus, especially peptic ulcer; (2) acute cholecystitis and biliary colic; (3) acute intestinal obstruction; (4) mesenteric vascular occlusion; (5) renal colic; (6) inferior myocardial infarction; (7) dissecting aortic aneurysm; (8) connective tissue disorders with vasculitis; (9) pneumonia; and (10) diabetic ketoacidosis. It may be difficult to differentiate acute cholecystitis from acute pancreatitis, because an elevated serum amylase may be found in both disorders. Pain of biliary tract origin is more right sided or epigastric than perumbilical or left upper quadrant and can be more severe; ileus is usually absent. Ultrasound is helpful in establishing the diagnosis of cholelithiasis and cholecystitis. Intestinal obstruction due to mechanical factors can be differentiated from pancreatitis by the history of crescendo-decrescendo pain, findings on abdominal examination, and CT of the abdomen showing changes characteristic of mechanical obstruction. Acute mesenteric vascular occlusion is usually suspected in elderly debilitated patients with leukocytosis, abdominal distention, and bloody diarrhea, confirmed by CT or magnetic resonance angiography. Vasculitides secondary to systemic lupus erythematosus and polyarteritis nodosa may be confused with pancreatitis, especially because pancreatitis may develop as a complication of these diseases. Diabetic ketoacidosis is often accompanied by abdominal pain and

TABLE 348-3 Severe Acute Pancreatitis

Risk Factors for Severity
<ul style="list-style-type: none"> • Age >60 years • Obesity, BMI >30 kg/m² • Comorbid disease (based on Charlson comorbidity index)
Markers of Severity at Admission or within 24 h
<ul style="list-style-type: none"> • SIRS—defined by presence of 2 or more criteria: <ul style="list-style-type: none"> • Core temperature <36° or >38°C • Heart rate >90 beats/min • Respirations >20/min or PaCO₂ <32 mmHg • White blood cell count >12,000/µL, <4000/µL, or 10% bands • APACHE II (≥8 at 24 h) • Hemoconcentration (hematocrit >44%) • Admission BUN (>22 mg/dL) • BISAP score (≥3 present) <ul style="list-style-type: none"> • (B) BUN >25 mg/dL • (I) Impaired mental status • (S) SIRS: ≥2 of 4 present • (A) Age >60 years • (P) Pleural effusion • Organ failure (Modified Marshall score) (≥1 present): <ul style="list-style-type: none"> • Cardiovascular: systolic BP <90 mmHg, heart rate >130 beats/min • Pulmonary: PaO₂ <60 mmHg • Renal: serum creatinine >2.0 mg/dL
Markers of Severity during Hospitalization
<ul style="list-style-type: none"> • Persistent organ failure (≥48 h) • Pancreatic or extrapancreatic necrosis

Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation II; BISAP, Bedside Index of Severity in Acute Pancreatitis; BMI, body mass index; BP, blood pressure; BUN, blood urea nitrogen; SIRS, systemic inflammatory response syndrome.

elevated total serum amylase levels, thus closely mimicking acute pancreatitis; however, the serum lipase level is often not elevated in diabetic ketoacidosis, and pancreas imaging is normal.

■ CLINICAL COURSE, DEFINITIONS, AND CLASSIFICATIONS

The Revised Atlanta Criteria define (1) phases of acute pancreatitis, (2) severity of acute pancreatitis, and (3) radiographic definitions, as outlined below.

Phases of Acute Pancreatitis Two phases of acute pancreatitis have been defined, early (<2 weeks) and late (>2 weeks), which primarily describe the hospital course of the disease. In the *early phase* of acute pancreatitis, which lasts 1–2 weeks, severity is defined by clinical parameters rather than morphologic findings. Most patients exhibit SIRS, and if this persists, patients are predisposed to organ failure. Three organ systems should be assessed to define organ failure: respiratory, cardiovascular, and renal. Organ failure is defined as a score of 2 or more for one of these three organ systems using the modified Marshall scoring system. Persistent organ failure (>48 h) is the most important clinical finding regarding severity of the acute pancreatitis episode. Organ failure that affects more than one organ is considered multisystem organ failure. CT imaging is usually not needed or recommended during the first 48 h of admission in acute pancreatitis.

The *late phase* is characterized by a protracted course of illness and may require imaging to evaluate for local complications. The critical clinical parameter of severity, as in the early phase, is persistent organ failure. These patients may require supportive measures such as renal dialysis, ventilator support, or need for supplemental nutrition via a nasojejunal or parenteral route. The radiographic feature of greatest importance to recognize in this phase is the development of necrotizing pancreatitis on CT imaging. Necrosis is associated with prolonged hospitalization and, if infected, may require intervention (percutaneous, endoscopic, and/or surgical).

Severity of Acute Pancreatitis Three classes of severity have been defined: mild, moderately severe, and severe. *Mild acute pancreatitis* is without local complications or organ failure. Most patients with interstitial acute pancreatitis have mild pancreatitis. In mild acute pancreatitis, the disease is self-limited and subsides spontaneously, usually within 3–7 days after onset. Oral intake can be resumed if the patient is hungry, has normal bowel function, and is without nausea and vomiting. Typically, a clear or full liquid diet has been recommended for the initial meal; however, a low-fat solid diet is a reasonable choice following recovery from mild acute pancreatitis.

Moderately severe acute pancreatitis is characterized by transient organ failure (i.e., it resolves in <48 h) or local or systemic complications in the absence of persistent organ failure. These patients may or may not have necrosis but may develop a local complication such as a fluid collection that requires a prolonged hospitalization >1 week. As with mild acute pancreatitis, the mortality rate for these patients remains low.

Severe acute pancreatitis is characterized by persistent organ failure (>48 h), involving one or more organs. A CT scan or magnetic resonance imaging (MRI) should be obtained to assess for necrosis and/or complications. If a local complication is encountered, management is dictated by clinical symptoms, evidence of infection, the maturity of fluid collection, and clinical stability of the patient. Prophylactic antibiotics are no longer recommended for severe acute pancreatitis.

Imaging in Acute Pancreatitis Two types of pancreatitis are recognized on imaging as *interstitial* or *necrotizing* based on pancreatic perfusion. CT imaging with IV contrast is best evaluated 3–5 days into hospitalization if patients are not responding to supportive care to assess for local complications such as necrosis. Recent studies report the overutilization of CT imaging within 72 h for acute pancreatitis, including those with a mild severity of disease. The Revised Atlanta Criteria also outline the terminology for local complications and fluid collections along with a CT imaging template to guide reporting of findings. Local morphologic features are summarized in Table 348-2. *Interstitial pancreatitis* occurs in 90–95% of admissions for acute pancreatitis and is characterized by diffuse gland enlargement, homogenous contrast enhancement, and mild inflammatory changes or peripancreatic stranding. Symptoms generally resolve with a week of hospitalization. *Necrotizing pancreatitis* occurs in 5–10% of acute pancreatitis admissions and may not evolve until several days of hospitalization. It is characterized by lack of pancreatic parenchymal enhancement by intravenous contrast agent and/or presence of findings of peripancreatic necrosis. The natural history of pancreatic and peripancreatic necrosis is variable because it may remain solid or liquefy, remain sterile or become infected, and persist or disappear over time. Importantly, those with only extrapancreatic necrosis have a more favorable prognosis than patients with pancreatic necrosis (with or without extrapancreatic necrosis). CT identification of local complications, particularly necrosis, is critical in patients who are not responding to therapy because patients with infected and sterile necrosis are at greatest risk of mortality (Figs. 348-1 and 348-2). The median prevalence of organ failure is >50% in necrotizing pancreatitis, and is perhaps slightly higher in infected versus sterile necrosis. With single-organ system failure, the mortality is 3–10%, but increases to nearly 50% with multiorgan failure.

■ ACUTE PANCREATITIS MANAGEMENT

The management of patients with acute pancreatitis from the time of diagnosis in the emergency ward to hospital discharge is briefly reviewed, highlighting salient features based on severity and complications. It is important to recognize that 85–90% of cases of acute pancreatitis are self-limited and subside spontaneously, usually within 3–7 days after onset, and do not exhibit organ failure or local complications.

The management of acute pancreatitis begins in the emergency ward. After a diagnosis has been confirmed, early and aggressive fluid resuscitation is critical. Additionally, intravenous analgesics are administered, severity is assessed, and a search for etiologies that may impact acute care is begun. Patients who do not respond to aggressive fluid

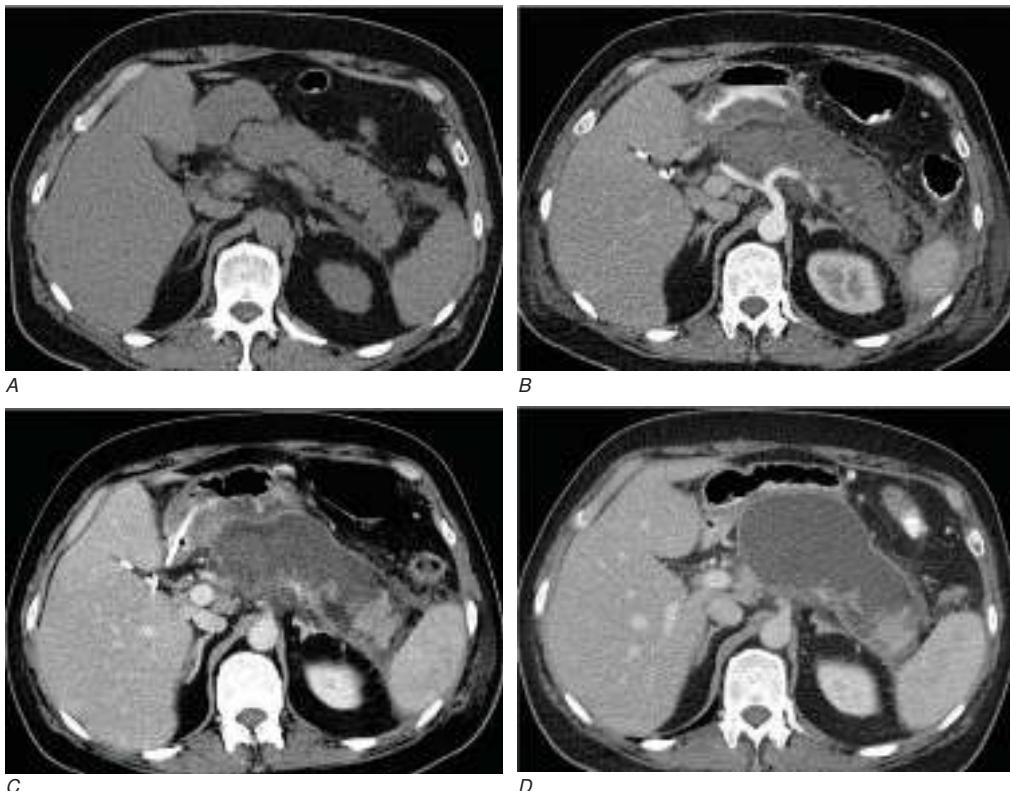


FIGURE 348-1 Evolution of changes of acute necrotizing pancreatitis on computed tomography (CT). **A.** CT scan of the abdomen without IV contrast performed on admission for a patient with acute gallstone pancreatitis, showing mild peripancreatic stranding. **B.** Contrast-enhanced CT scan of the abdomen performed on the same patient 1 week after admission shows extensive intrapancreatic necrosis, evidenced by the lack of contrast enhancement in the pancreatic body with very minimal enhancement noted at the distal most aspect of the pancreatic tail. **C.** Contrast-enhanced CT scan of the abdomen performed on the same patient 2 weeks after admission demonstrates a semiorganized, heterogeneous fluid collection, referred to as an acute necrotic collection. On this image, a small area of viable pancreatic parenchyma is seen at the tail of the pancreas. **D.** Contrast-enhanced CT scan of the abdomen performed on the same patient 5 weeks after admission demonstrates a well-encapsulated fluid collection, essentially replacing the pancreas, referred to as walled-off necrosis.

resuscitation in the emergency ward should be considered for admission to a step-down or intensive care unit for aggressive fluid resuscitation, hemodynamic monitoring, and management of any organ failure.

Fluid Resuscitation and Monitoring Response to Therapy The most important treatment intervention for acute pancreatitis is early, aggressive intravenous fluid resuscitation to prevent systemic complications from the secondary systemic inflammatory response. The patient is initially made NPO to minimize nutrient-induced stimulation of

the pancreas and is given intravenous narcotic analgesics to control abdominal pain and supplemental oxygen (as needed).

Intravenous fluids of lactated Ringer's or normal saline are initially bolused at 15–20 mL/kg (1050–1400 mL), followed by 2–3 mL/kg per hour (200–250 mL/h), to maintain urine output >0.5 mL/kg per hour. Serial bedside evaluations are required every 6–8 h to assess vital signs, oxygen saturation, and change in physical examination to optimize fluid resuscitation. Lactated Ringer's solution has been shown to



FIGURE 348-2 Imaging features of a pancreaticopleural fistula secondary to acute pancreatitis. **A.** Pancreaticopleural fistula: pancreatic duct leak on endoscopic retrograde cholangiopancreatography. Pancreatic duct leak (arrow) demonstrated at the time of retrograde pancreateogram in a patient with exacerbation of alcohol-induced acute pancreatitis. **B.** Pancreaticopleural fistula: computed tomography (CT) scan. Contrast-enhanced CT scan (coronal view) with arrows showing fistula tract from pancreatic duct disruption in the pancreatic pleural fistula. **C.** Pancreaticopleural fistula: chest x-ray. Large pleural effusion in the left hemithorax from a disrupted pancreatic duct. Analysis of pleural fluid revealed elevated amylase concentration. (Courtesy of Dr. K.J. Morteze, Brigham and Women's Hospital; with permission.)

decrease systemic inflammation (lower C-reactive protein levels from admission) and may be a better crystalloid than normal saline. A targeted resuscitation strategy with measurement of hematocrit and BUN every 8–12 h is recommended to ensure adequacy of fluid resuscitation and monitor response to therapy, noting that a less aggressive resuscitation strategy may be needed in milder forms of pancreatitis. A rising BUN during hospitalization is not only associated with inadequate hydration but also higher in-hospital mortality.

A decrease in hematocrit and BUN during the first 12–24 h is strong evidence that sufficient fluids are being administered. Serial measurements and bedside assessment for fluid overload are continued, and fluid rates are maintained at the current rate. Adjustments in fluid resuscitation may be required in patients with cardiac, pulmonary, or renal disease. A rise in hematocrit or BUN during serial measurement should be treated with a repeat volume challenge with a 2-L crystalloid bolus followed by increasing the fluid rate by 1.5 mg/kg per hour. If the BUN or hematocrit fails to respond (i.e., remains elevated or does not decrease) to this bolus challenge and increase in fluid rate, consideration of transfer to an intensive care unit is strongly recommended for hemodynamic monitoring.

Assessment of Severity and Hospital Triage Severity of acute pancreatitis should be determined in the emergency ward to assist in patient triage to a regular hospital ward or step-down unit or direct admission to an intensive care unit. The Bedside Index of Severity in Acute Pancreatitis (BISAP) incorporates five clinical and laboratory parameters obtained within the first 24 h of hospitalization (Table 348-3)—BUN >25 mg/dL, impaired mental status (Glasgow coma scale score <15), SIRS, age >60 years, and pleural effusion on radiography—that can be useful in assessing severity. The presence of three or more of these factors was associated with substantially increased risk for in-hospital mortality among patients with acute pancreatitis. In addition, an elevated hematocrit >44% and admission BUN >22 mg/dL are also associated with more severe acute pancreatitis. Incorporating these indices with the overall patient response to initial fluid resuscitation in the emergency ward can be useful at triaging patients to the appropriate hospital acute care setting.

In general, patients with lower BISAP scores, hematocrits, and admission BUNs tend to respond to initial management and can be safely triaged to a regular hospital ward for ongoing care. If SIRS is not present at 24 h, the patient is unlikely to develop organ failure or necrosis. Therefore, patients with persistent SIRS at 24 h or underlying comorbid illnesses (e.g., chronic obstructive pulmonary disease, congestive heart failure) should be considered for a step-down unit setting if available. Patients with higher BISAP scores and elevations in hematocrit and admission BUN who do not respond to initial fluid resuscitation and exhibit evidence of respiratory failure, hypotension, or organ failure should be considered for direct admission to an intensive care unit.

Special Considerations Based on Etiology A careful history, review of medications, selected laboratory studies (liver profile, serum triglycerides, serum calcium), and an abdominal ultrasound are recommended in the emergency ward to assess for etiologies that may impact acute management. An abdominal ultrasound is the initial imaging modality of choice and will evaluate the gallbladder, common bile duct, and pancreatic head.

GALLSTONE PANCREATITIS Patients with evidence of ascending cholangitis (rising white blood cell count, increasing liver enzymes) should undergo ERCP within 24–48 h of admission. Patients with gallstone pancreatitis are at increased risk of recurrence, and consideration should be given to performing a cholecystectomy during the same admission in mild acute pancreatitis. An alternative for patients who are not surgical candidates would be to perform an endoscopic biliary sphincterotomy before discharge.

HYPERTRIGLYCERIDEMIA Serum triglycerides >1000 mg/dL are associated with acute pancreatitis. Initial therapy should focus on treatment of hyperglycemia with intravenous insulin, which often corrects the hypertriglyceridemia. Adjunct therapies may also include heparin or

plasmapheresis, but there is no compelling evidence these measures improve clinical outcomes. Outpatient therapies include control of diabetes if present, administration of lipid-lowering agents, weight loss, and avoidance of drugs that elevate lipid levels.

Other potential etiologies that may impact acute hospital care include *hypercalcemia* and *post-ERCP pancreatitis*. Treatment of hyperparathyroidism or malignancy is effective at reducing serum calcium. Pancreatic duct stenting and rectal indomethacin administration are effective at decreasing pancreatitis after ERCP. Drugs that cause pancreatitis should be discontinued. Multiple drugs have been implicated, but only about 30 have been rechallenged (Class 1A) and found to be causative.

Nutritional Therapy A low-fat solid diet can be administered to subjects with mild acute pancreatitis once they are able to eat. Enteral nutrition should be considered 2–3 days after admission in subjects with more severe pancreatitis instead of total parenteral nutrition (TPN). Enteral feeding maintains gut barrier integrity, limits bacterial translocation, is less expensive, and has fewer complications than TPN. Gastric feeding is safe; the benefits of nasojejunal enteral feeding over gastric feeding remains under investigation.

Management of Local Complications (Table 348-4) Patients exhibiting signs of clinical deterioration despite aggressive fluid resuscitation and hemodynamic monitoring should be assessed for local complications, which may include necrosis, pseudocyst formation, pancreas duct disruption, peripancreatic vascular complications, and extrapancreatic infections. A multidisciplinary team approach is recommended, including gastroenterology, surgery, interventional radiology, and intensive care specialists, and consideration should also be made for transfer to a tertiary pancreas center of excellence.

NECROSIS The management of necrosis requires a multidisciplinary team approach. Percutaneous fine-needle aspiration of necrosis with Gram stain and culture was previously performed to evaluate for infected pancreatic necrosis in those with sustained leukocytosis, fever, or organ failure. However, the current use of this technique varies depending on institutional preference, with many abandoning this diagnostic test to avoid potentially contaminating an otherwise sterile collection, particularly when culture results will not lead to a clinical decision to de-escalate antimicrobial therapy. Even though there is currently no role for *prophylactic antibiotics* in necrotizing pancreatitis, *empiric antibiotics* should be considered in those with clinical decompensation. Prophylactic antibiotics do not lead to improved survival and may promote the development of opportunistic fungal infections. Repeated CT or MRI imaging should also be considered with any change in clinical course to monitor for complications (e.g., thrombosis, hemorrhage, abdominal compartment syndrome).

In general, *sterile necrosis* is most often managed conservatively unless complications arise. Once a diagnosis of *infected necrosis* is established and an organism identified, targeted antibiotics should be instituted. Pancreatic drainage and/or debridement (necrosectomy) should be considered for definitive management of *infected necrosis*, but clinical decisions are ultimately influenced by the clinical response since almost two-thirds of patients respond to antibiotic treatment with or without percutaneous drainage. Symptomatic local complications as outlined in the Revised Atlanta Criteria typically require definitive therapy.

A step-up approach (percutaneous or endoscopic transgastric/transduodenal drainage followed, if necessary, by surgical necrosectomy) has been successfully reported by some pancreatic centers. One-third of the patients successfully treated with the step-up approach did not require major abdominal surgery. A randomized trial reported advantages to an initial endoscopic approach compared to an initial surgical necrosectomy approach in select patients requiring intervention for symptomatic WON. Taken together, a more conservative approach to the management of infected pancreatic necrosis has evolved under the close supervision of a multidisciplinary team. If conservative therapy can be safely implemented, it is recommended to do so for 4–6 weeks to allow the pancreatic collections to either resolve or evolve to develop

TABLE 348-4 Complications of Acute Pancreatitis

Local

Pancreatic/peripancreatic fluid collections:

- Acute necrotic collection (sterile or infected)
- Walled-off necrosis (sterile or infected)
- Pancreatic pseudocyst

Disruption of main pancreatic duct or secondary branches

Pancreatic ascites

Chylous ascites (secondary to disruption of lymphatic ducts)

Involvement of contiguous organs by necrotizing pancreatitis (e.g., colon perforation)

Splanchnic thromboses (splenic vein, superior mesenteric vein, and/or portal vein)

Bowel infarction/perforation

Gastric outlet obstruction

Biliary obstruction (jaundice)

Systemic

Pulmonary

- Pleural effusion
- Atelectasis
- Mediastinal fluid
- Pneumonitis
- Acute respiratory distress syndrome
- Hypoxemia (unrecognized)

Cardiovascular

- Hypotension
- Hypovolemia
- Nonspecific ST-T changes in electrocardiogram simulating myocardial infarction
- Pericardial effusion

Hematologic

- Disseminated intravascular coagulation

Gastrointestinal hemorrhage

- Peptic ulcer disease
- Erosive gastritis
- Hemorrhagic pancreatic necrosis with erosion into major blood vessels
- Variceal hemorrhage secondary to splanchnic thrombosis

Renal

- Oliguria (<300 mL/d)
- Azotemia
- Renal artery and/or renal vein thrombosis
- Acute tubular necrosis

Metabolic

- Hyperglycemia
- Hypertriglyceridemia
- Hypocalcemia
- Encephalopathy
- Sudden blindness (Purtscher's retinopathy)

Central nervous system

- Psychosis
- Fat emboli
- Fat necrosis
- Subcutaneous tissues (erythematous nodules)
- Bone
- Miscellaneous (mediastinum, pleura, nervous system)

a more organized boundary (i.e., to "wall off") so that surgical or endoscopic intervention is generally safer and more effective.

PSEUDOCYST The incidence of pseudocyst is low, and most acute collections resolve over time. Less than 10% of patients have persistent fluid collections after 4 weeks that would meet the definition of a pseudocyst. Only symptomatic collections require intervention with endoscopic or surgical drainage.

PANCREATIC DUCT DISRUPTION Pancreatic duct disruption may present with symptoms of increasing abdominal pain or shortness of breath in the setting of an enlarging fluid collection resulting in pancreatic ascites (ascitic fluid has high amylase level). Diagnosis can be confirmed on magnetic resonance cholangiopancreatography (MRCP) or ERCP. Placement of a bridging pancreatic stent for at least 6 weeks is >90% effective at resolving the leak with or without parenteral nutrition and octreotide. Nonbridging stents are less effective (25–50%) but should be considered with parenteral nutrition and octreotide prior to surgical intervention.

PERIVASCULAR COMPLICATIONS Perivascular complications may include *splenic vein thrombosis* with gastric varices and pseudoaneurysms, as well as *portal and superior mesenteric vein thromboses*. *Gastric varices* rarely bleed but can be life-threatening. Similarly, life-threatening bleeding from a ruptured *pseudoaneurysm* can be diagnosed and treated with mesenteric angiography and embolization.

EXTRAPANCREATIC INFECTIONS Hospital-acquired infections occur in up to 20% of patients with acute pancreatitis. Patients should be continually monitored for the development of pneumonia, urinary tract infection, and line infection. Continued culturing of urine, monitoring of chest x-rays, and routine changing of intravenous lines are important during hospitalization.

Follow-Up Care Hospitalizations for moderately severe and severe acute pancreatitis can be prolonged and last weeks to months and often involve periods of intensive care unit admission and outpatient rehabilitation or subacute nursing care. Follow-up evaluation should assess for development of diabetes, exocrine pancreatic insufficiency, recurrent cholangitis, or infected fluid collections. As mentioned previously, cholecystectomy should be performed during the initial hospitalization for acute gallstone pancreatitis with mild clinical severity. For patients with necrotizing gallstone pancreatitis, the timing of cholecystectomy needs to be individualized.

■ RECURRENT ACUTE PANCREATITIS

Approximately 25% of patients who have had an attack of acute pancreatitis have a recurrence. The two most common etiologic factors are alcohol and cholelithiasis. In patients with recurrent pancreatitis without an obvious cause, the differential diagnosis should encompass occult biliary tract disease, including microlithiasis, hypertriglyceridemia, pancreatic cancer, and hereditary pancreatitis (Table 348-1). In one series of 31 patients diagnosed initially as having idiopathic or recurrent acute pancreatitis, 23 were found to have occult gallstone disease. Thus, approximately two-thirds of patients with recurrent acute pancreatitis without an obvious cause actually have occult gallstone disease due to microlithiasis. Genetic defects as in hereditary pancreatitis and cystic fibrosis mutations can result in recurrent pancreatitis. Other diseases of the biliary tree and pancreatic ducts that can cause acute pancreatitis include choledochocoele; ampullary tumors; pancreas divisum; and pancreatic duct stones, stricture, and tumor. Approximately 2–4% of patients with pancreatic cancer present with acute pancreatitis.

■ PANCREATITIS IN PATIENTS WITH AIDS

The incidence of acute pancreatitis is theoretically increased in patients with AIDS for two reasons: (1) the high incidence of infections involving the pancreas such as infections with cytomegalovirus, *Cryptosporidium*, and the *Mycobacterium avium* complex; and (2) the frequent use by patients with AIDS of medications such as pentamidine, trimethoprim-sulfamethoxazole, and protease inhibitors. The incidence has been markedly reduced due to advances in therapy, including the disuse of didanosine (Chap. 202).

CHRONIC PANCREATITIS AND EXOCRINE PANCREATIC INSUFFICIENCY

■ PATHOPHYSIOLOGY

Chronic pancreatitis is a disease process characterized by irreversible damage to the pancreas, in contrast to the reversible changes noted

in acute pancreatitis (Table 348-4). The events that initiate and then perpetuate the inflammatory process in the pancreas are becoming more clearly understood. Irrespective of the mechanism of injury, it is becoming apparent that stellate cell activation leads to cytokine expression and production of extracellular matrix proteins that contribute to acute and chronic inflammation and collagen deposition in the pancreas. This condition is defined by the presence of histologic abnormalities, including chronic inflammation, fibrosis, and progressive destruction (atrophy) of both exocrine and endocrine tissue. A number of etiologies have been associated with chronic pancreatitis resulting in the cardinal manifestations of the disease such as abdominal pain, steatorrhea, weight loss, diabetes mellitus, and, less commonly, pancreatic cancer (Table 348-5).

Even in individuals in whom alcohol is believed to be the primary cause of chronic pancreatitis, other factors are likely required for the development and progression of disease, which explains why not all heavy consumers of alcohol develop pancreatic disease. There is also a strong association between smoking and chronic pancreatitis. Cigarette smoke leads to an increased susceptibility to pancreatic autodigestion and predisposes to dysregulation of duct cell CFTR function. Smoking is an independent, dose-dependent risk factor for chronic pancreatitis and recurrent acute pancreatitis. Both continued alcohol and smoking exposure are associated with disease progression, including pancreatic fibrosis and calcifications.

Characterization of pancreatic stellate cells (PSCs) has added insight into the underlying cellular responses behind development of chronic pancreatitis. Specifically, PSCs are believed to play a role in maintaining

normal pancreatic architecture that shifts toward fibrogenesis in those who develop chronic pancreatitis. It is believed that alcohol or additional stimuli lead to matrix metalloproteinase-mediated destruction of normal collagen in pancreatic parenchyma, which later allows for pancreatic remodeling. Proinflammatory cytokines, tumor necrosis factor α (TNF- α), interleukin 1 (IL-1), and interleukin 6 (IL-6), as well as oxidant complexes, can induce PSC activity with subsequent new collagen synthesis. In addition to being stimulated by cytokines, oxidants, or growth factors, PSCs also possess transforming growth factor β (TGF- β)-mediated self-activating autocrine pathways that may explain disease progression in chronic pancreatitis even after removal of noxious stimuli.

■ ETIOLOGIC CONSIDERATIONS

Among adults in the United States, alcoholism is the most common cause of clinically apparent chronic pancreatitis, whereas cystic fibrosis is the most frequent cause in children. As many as 25% of adults in the United States with chronic pancreatitis have the *idiopathic* form, including a subset of patients who do not develop clinical manifestations until later in life (*idiopathic late-onset chronic pancreatitis*). Recent investigations have indicated that up to 15% of patients with chronic pancreatitis previously classified as having idiopathic pancreatitis may have an underlying genetic predisposition (Table 348-5).

The prototypical genetic defect was identified in the cationic trypsinogen gene (*PRSS1*) by studying several large families with chronic pancreatitis. Additional pathogenic and nonpathogenic mutations have been identified in this gene. The defect prevents the destruction of prematurely activated trypsin and allows it to be resistant to the intracellular protective effect of trypsin inhibitor. It is hypothesized that this continual activation of digestive enzymes within the gland leads to acute injury and, finally, chronic pancreatitis. Since the initial discovery of the *PRSS1* mutation defect, other genetic disease modifiers have been identified (Table 348-5).

The *CFTR* gene functions as a cyclic AMP-regulated chloride channel. In patients with cystic fibrosis, the high concentration of macromolecules can block the pancreatic ducts. It must be appreciated, however, that there is a great deal of heterogeneity in relationship to the *CFTR* gene defect. More than 1700 putative mutations of the *CFTR* gene have been identified. Attempts to elucidate the relationship between the genotype and pancreatic manifestations have been hampered by the large number and different classes of *CFTR* mutations. The ability to detect *CFTR* mutations has led to the recognition that the clinical spectrum of the disease is broader than previously thought. Two studies have clarified the association between mutations of the *CFTR* gene and another monosymptomatic form of cystic fibrosis (i.e., chronic pancreatitis). It is estimated that in patients with idiopathic pancreatitis, the frequency of a single *CFTR* mutation is 11 times the expected frequency and the frequency of two mutant alleles is 80 times the expected frequency. In these studies, patients were adults when the diagnosis of pancreatitis was made; none had any clinical evidence of pulmonary disease, and sweat test results were not diagnostic of cystic fibrosis. The prevalence of such mutations is unclear, and further studies are needed. In addition, the therapeutic and prognostic implication of these findings with respect to managing pancreatitis remains to be determined. *CFTR* mutations are common in the general population, so it is unclear whether the *CFTR* mutation alone can lead to pancreatitis as an autosomal recessive disease. A study evaluated 39 patients with idiopathic chronic pancreatitis to assess the risk associated with these mutations. Patients with two *CFTR* mutations (compound heterozygotes) demonstrated *CFTR* function at a level between that seen in typical cystic fibrosis and cystic fibrosis carriers and had a fortyfold increased risk of pancreatitis. The presence of a separate genetic mutation (*N34S SPINK1*) increased the risk twentyfold. A combination of two *CFTR* mutations and an *N34S SPINK1* mutation increased the risk of pancreatitis 900-fold. Knowledge of the genetic defects and downstream alterations in protein expression has led to the development of novel therapies in children with cystic fibrosis that potentiate the *CFTR* channel, resulting in improvement in lung function, quality of life, and weight gain. Some studies have shown that use of *CFTR* modulators

TABLE 348-5 Classification of Chronic Pancreatitis: The TIGAR-O System

Toxic-metabolic
Alcoholic
Tobacco smoking
Hypercalcemia
Hyperlipidemia (hypertriglyceridemia)
Chronic renal failure
Idiopathic
Early onset
Late onset
Tropical
Genetic
Cationic trypsinogen (<i>PRSS1</i>)
Cystic fibrosis transmembrane conductance regulator gene (<i>CFTR</i>) ^a
Calcium-sensing receptor (<i>CASR</i>) ^a
Chymotrypsin C gene (<i>CTRC</i>) ^a
Pancreatic secretory trypsin inhibitor gene (<i>SPINK1</i>) ^a
Autoimmune
Type 1 autoimmune pancreatitis (associated with IgG4-related disease)
Type 2 autoimmune pancreatitis (idiopathic duct-centric chronic pancreatitis)
Recurrent and severe acute pancreatitis
Postnecrotic (severe acute pancreatitis)
Recurrent acute pancreatitis
Vascular diseases/ischemia
Radiation induced
Obstructive
Pancreas divisum ^a
Duct obstruction (e.g., tumor)
Preampillary duodenal wall cysts
Posttraumatic pancreatic duct strictures

^aThese conditions are believed to be disease modifiers that require additional factors to cause chronic pancreatitis.

Abbreviations: TIGAR-O, toxic-metabolic, idiopathic, genetic, autoimmune, recurrent and severe acute pancreatitis, obstructive.

TABLE 348-6 Comparison of the Autoimmune Pancreatitis (AIP) Subtypes

	TYPE 1 AIP	TYPE 2 AIP
Age at diagnosis, mean	Seventh decade	Fifth decade
Male sex	75%	50%
Serum IgG4 elevation	~66%	~25%
Other organ involvement	50%	No ^a
Histologic findings:		
Lymphoplasmacytic infiltration	++	++
Periductal inflammation	++	++
Storiform fibrosis	++	+
Obliterative phlebitis	++	+
Granulocytic epithelial lesion (GEL)	-	+++
IgG4 tissue staining	Abundant (≥10 cells/hpf)	Scant (<10 cells/hpf)
Response to steroids	~100%	~100%
Risk for relapse	High (20–60%)	Low (<10%)
Associated with IgG4-RD	Yes	No

^aInflammatory bowel disease is seen in ~10–20% of patients with idiopathic duct-centric chronic pancreatitis but may also occur in type 1 AIP.

Abbreviation: IgG4-RD, IgG4-related disease.

Source: Reproduced with permission from PA Hart: Reviews in basic and clinical gastroenterology and hepatology. Gastroenterology 149:39, 2015.

may reduce the frequency of acute pancreatitis in heterozygous carriers. Table 348-5 lists other recognized causes of chronic pancreatitis.

AUTOIMMUNE PANCREATITIS **TABLE 348-6**

Autoimmune pancreatitis (AIP) refers to a form of chronic pancreatitis with distinct histopathology and several unique differences in the clinical phenotype. Currently, two subtypes of AIP are recognized, type 1 AIP and idiopathic duct-centric chronic pancreatitis (IDCP, also referred to as type 2 AIP). Type 1 AIP is identified as the pancreatic manifestation of a multiorgan syndrome currently referred to as IgG4-related disease (Chap. 368). The characteristic histopathologic findings of type 1 AIP include lymphoplasmacytic infiltrate, storiform fibrosis, and abundant IgG4 cells. IDCP is histologically defined by the presence of granulocytic infiltration of the duct wall (termed a granulocytic epithelial lesion [GEL]) but without IgG4-positive cells. Type 1 AIP is often associated with involvement of other organs in the setting of IgG4-related disease, including bilateral submandibular gland enlargement, characteristic renal lesions, retroperitoneal fibrosis, and stricturing of the suprapancreatic biliary tree. In contrast, IDCP is

a pancreas-specific disorder that is associated with inflammatory bowel disease in ~10% of patients. AIP is not a common cause of idiopathic recurrent acute pancreatitis.

Jaundice, weight loss, and new-onset diabetes are the most common presenting symptoms. Elevated serum IgG4 levels are supportive of the diagnosis (elevated in two-thirds of patients with type 1 AIP) but have a low positive predictive value when used in isolation of other clinical findings. CT imaging demonstrates abnormalities in the majority of patients with either diffuse or focal enlargement during active disease, unless the gland is atrophic due to previous disease (Fig. 348-3). The presence of an inflammatory rim, termed a capsule sign, is highly specific (but not sensitive) for AIP. ERCP or MRCP reveals strictures in the bile duct in more than one-third of patients with AIP, including some patients with isolated intrahepatic bile duct strictures (type 1 AIP only), which can mimic primary sclerosing cholangitis, and is referred to as IgG4-related sclerosing cholangitis (previously termed IgG4-associated cholangitis).

The Mayo Clinic HISORt criteria provide a helpful mnemonic to remember the key diagnostic features of this disease, including (1) histology; (2) imaging; (3) serology (elevated serum IgG4 levels); (4) other organ involvement; and (5) response to glucocorticoid therapy. These diagnostic criteria have been harmonized with those from other countries to develop the International Consensus Diagnostic Criteria for AIP, which are the most widely utilized criteria. Glucocorticoids have shown efficacy in alleviating symptoms, decreasing the size of the pancreas, and reversing histopathologic features in patients with AIP. Patients typically respond dramatically to glucocorticoid therapy within a 2- to 4-week period. Prednisone is usually administered at an initial dose of 40 mg/d for 4 weeks followed by a taper of the daily dosage by 5 mg per week based on monitoring of clinical parameters. Relief of symptoms, liver biochemistries, and abnormal imaging of the pancreas and bile ducts are followed to assess for treatment response. A poor response to glucocorticoids should raise suspicion of an alternate diagnosis, such as pancreatic cancer. A recent multicenter international study examined >1000 patients with AIP. Clinical remission was achieved in 99% of type 1 AIP and 92% of type 2 AIP patients with steroids. However, disease relapse occurred in 31 and 9% of patients with type 1 and type 2 AIP, respectively. Patients with multiple relapses may be managed with an immunomodulator (e.g., azathioprine, 6-mercaptopurine, or mycophenolate mofetil) or B-cell depletion therapy (e.g., rituximab). The appearance of interval cancers following a diagnosis of AIP is uncommon.

Clinical Features of Chronic Pancreatitis Patients with chronic pancreatitis primarily seek medical attention due to abdominal pain or symptoms of maldigestion. The abdominal pain may be quite variable in location, severity, and frequency. The pain can be constant or

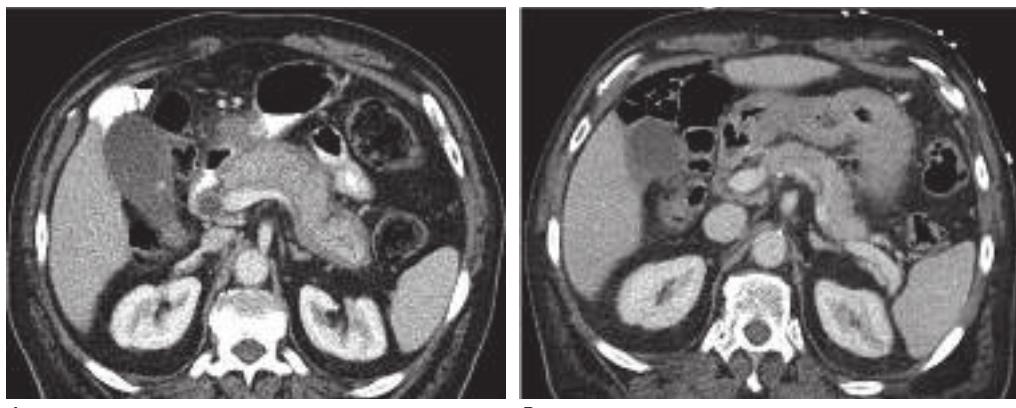


FIGURE 348-3 Imaging features of the pancreatic parenchyma in a patient with type 1 autoimmune pancreatitis on computed tomography (CT). *A*, Contrast-enhanced CT scan of the abdomen demonstrates diffuse pancreatic enlargement and a hypodense rim (capsule sign) in a patient who presented with jaundice. The serum IgG4 level was elevated to 942 mg/dL (reference range 4–86 mg/dL), so the patient was diagnosed with definitive type 1 autoimmune pancreatitis. *B*, Contrast-enhanced CT scan of the abdomen following a treatment course with high-dose steroids demonstrates return to normal size of the pancreas, reappearance of normal lobulations along the margin, and absence of the hypodense rim.

intermittent with pain-free intervals. Eating may exacerbate the pain, leading to a fear of eating with consequent weight loss. The spectrum of abdominal pain ranges from mild to quite severe, with narcotic dependence as a frequent consequence. There is often a disparity between the reported severity of abdominal pain and the physical findings, which primarily consist of nonfocal abdominal tenderness. Patients with chronic abdominal pain may or may not experience symptoms of maldigestion, such as chronic diarrhea, steatorrhea, and/or weight loss. Fat-soluble vitamin deficiencies are increasingly recognized. Importantly, there is an exceedingly high prevalence of metabolic bone disease in chronic pancreatitis, with ~65% of patients having either osteopenia or osteoporosis. Patients with chronic pancreatitis have impaired quality of life and develop significant morbidity, requiring frequent use of health care resources.

The diagnosis of early or mild chronic pancreatitis can be challenging because there is no accurate biomarker for the disease. In contrast to acute pancreatitis, the serum amylase and lipase levels are usually not strikingly elevated in chronic pancreatitis. Rather, low serum pancreatic enzyme levels are moderately specific for a diagnosis of chronic pancreatitis but have poor sensitivity. Elevation of serum bilirubin and alkaline phosphatase may indicate cholestasis secondary to common bile duct stricture caused by chronic inflammation or fibrosis. The cumulative prevalence of exocrine pancreatic insufficiency is >80%. The presence of overt steatorrhea in a patient with chronic pancreatitis is highly suggestive of this complication. However, in those with milder symptoms, additional testing, such as a random fecal elastase-1 level (on a formed stool specimen) may be needed to confirm the diagnosis of exocrine pancreatic insufficiency. The radiographic evaluation of a patient with suspected chronic pancreatitis usually proceeds from a noninvasive to more invasive approach. Abdominal CT

imaging (Fig. 348-4) is the initial modality of choice, followed by MRI, endoscopic ultrasound, and pancreas function testing. In addition to excluding a pseudocyst and pancreatic cancer, CT imaging may show calcifications, dilated pancreatic or biliary ducts, or an atrophic pancreas. Although abdominal CT scanning and MRCP greatly aid in the diagnosis of pancreatic disease, the diagnostic test with the best sensitivity is the hormone stimulation test using secretin. The secretin test becomes abnormal when $\geq 50\%$ of the pancreatic exocrine function has been lost. This usually correlates well with the onset of chronic abdominal pain. The role of endoscopic ultrasonography (EUS) in diagnosing early chronic pancreatitis is still evolving. A total of nine endoscopic features have been described in chronic pancreatitis. The presence of five or more features is considered diagnostic of chronic pancreatitis. EUS is not a specific enough test for detecting early chronic pancreatitis alone (Chap. 347) and may show positive features in patients with diabetes, patients with a history of cigarette smoking, or even in normal aging individuals. Recent data suggest that EUS can be combined with endoscopic pancreatic function testing (EUS-ePFT) during a single endoscopy to screen for chronic pancreatitis in patients with chronic abdominal pain. Diffuse calcifications noted on plain film of the abdomen usually indicate significant damage to the pancreas and are pathognomonic for chronic pancreatitis. Although alcohol is by far the most common cause of pancreatic calcification, such calcification may also be noted in hereditary pancreatitis, posttraumatic pancreatitis, idiopathic chronic pancreatitis, and tropical pancreatitis.

Complications of Chronic Pancreatitis There are a number of disease-related complications from chronic pancreatitis in addition to the aforementioned abdominal pain and exocrine pancreatic insufficiency (Table 348-7). The lifetime prevalence of chronic

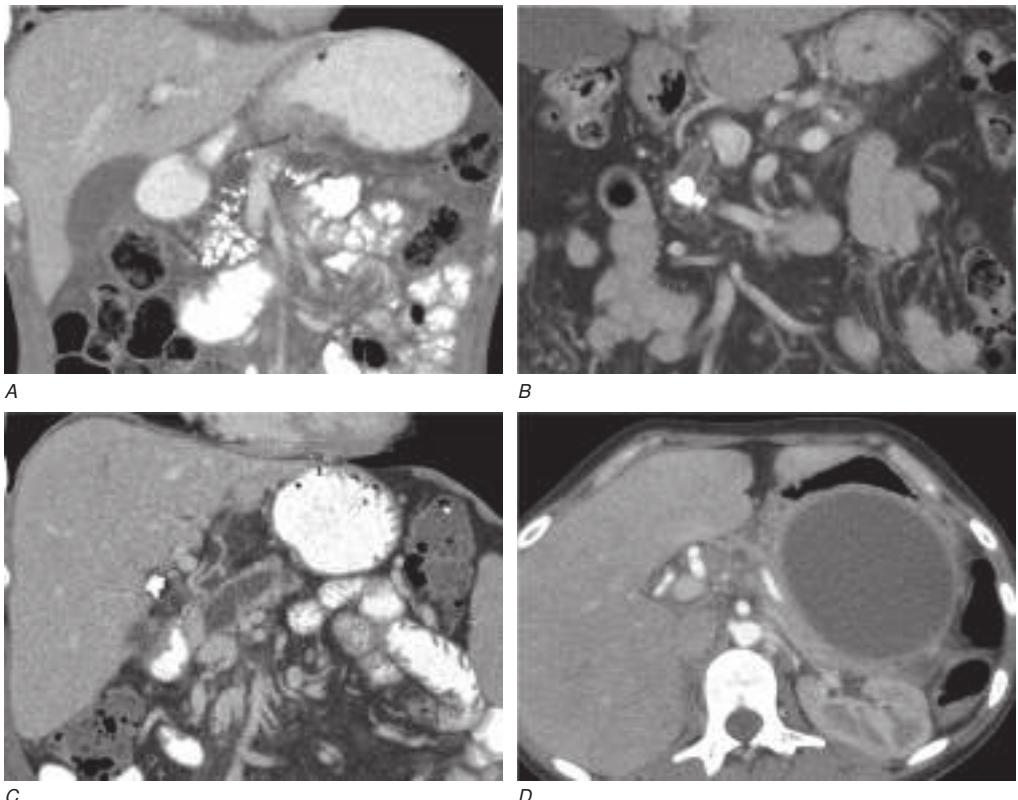


FIGURE 348-4 Distribution of imaging features of chronic pancreatitis on computed tomography (CT). Distinct features of chronic pancreatitis are seen on selected images from contrast-enhanced CT scans of the abdomen from four unique patients, including the following. *A*. Numerous punctate calcifications involving the pancreatic parenchyma in the head and body. *B*. A moderate-sized calculus visualized in the pancreatic duct with associated ductal dilation. *C*. Significant pancreatic duct dilation and adjacent parenchymal atrophy secondary to a pancreatic duct stricture (which is not well seen on this scan). *D*. A large unilocular, encapsulated cyst in the tail of the pancreas consistent with a pseudocyst from prior pancreatitis. Note adjacent pancreatic parenchymal atrophy.

TABLE 348-7 Complications of Chronic Pancreatitis

Chronic abdominal pain	Biliary stricture and/or biliary cirrhosis
Exocrine pancreatic insufficiency	Pancreatic duct stricture
Diabetes mellitus	Pseudocyst
Splanchnic venous thrombosis	Pancreatic cancer
Metabolic bone disease (osteoporosis)	

pancreatitis-related diabetes exceeds 80%. Although most patients develop hyperglycemia due to insulin deficiency caused by loss of islet cells, diabetic ketoacidosis and diabetic coma are uncommon. Likewise, end-organ damage (retinopathy, neuropathy, nephropathy) is also uncommon. Nondiabetic retinopathy may be due to vitamin A and/or zinc deficiency. Osteoporosis and osteopenia are increasingly recognized in chronic pancreatitis and likely related to a combination of shared risk factors (e.g., alcohol use, cigarette smoking), vitamin D deficiency, and detrimental effects on the bone from chronic inflammation. Gastrointestinal bleeding may occur from peptic ulceration, gastritis, a pseudocyst eroding into the duodenum, arterial bleeding into the pancreatic duct (hemosuccus pancreaticus), or ruptured varices secondary to splenic vein thrombosis. Jaundice, cholestasis, and biliary cirrhosis may occur from the chronic inflammatory reaction around the intrapancreatic portion of the common bile duct. Twenty years after the diagnosis of chronic calcific pancreatitis, the cumulative risk of pancreatic cancer is 4%. Patients with hereditary *PRSS1* or tropical pancreatitis have an increased risk for pancreatic cancer compared to other forms of chronic pancreatitis.

TREATMENT

Chronic Pancreatitis

There are currently no therapies to reverse or delay the disease progression of chronic pancreatitis, so management is primarily focused on screening for and management of disease-related complications.

STEATORRHEA

The treatment of steatorrhea with pancreatic enzyme replacement therapy is conceptually straightforward, yet complete correction of steatorrhea is uncommon. Enzyme therapy usually brings diarrhea under control and restores absorption of fat to an acceptable level and affects weight gain. Thus, pancreatic enzyme replacement is the cornerstone of therapy. In treating steatorrhea, it is important to use a potent pancreatic formulation that will deliver sufficient lipase into the duodenum to correct maldigestion and decrease steatorrhea. For adult patients with exocrine pancreatic insufficiency, it is generally recommended to start at a dosage of 25,000–50,000 units of lipase taken during each meal; however, the dose may need to be increased up to 100,000 units of lipase depending on the response in symptoms, nutritional parameters, and/or pancreas function test results. Additionally, some may require acid suppression with proton pump inhibitors to optimize the response to pancreatic enzymes. Monitoring nutritional parameters such as fat-soluble vitamins, zinc levels, body weight, and periodic bone mineral density measurement should be considered.

ABDOMINAL PAIN

The management of pain in patients with chronic pancreatitis is challenging due to the complex mechanisms of pancreatitis-related pain. Recent meta-analyses have shown no consistent benefit of enzyme therapy at reducing pain in chronic pancreatitis. Pain relief experienced by patients treated with pancreatic enzymes may be due to improvements in the dyspepsia from maldigestion. One short-term randomized trial showed that pregabalin could decrease pain in chronic pancreatitis and lower pain medication requirement. Other studies using antioxidants have yielded mixed results.

Endoscopic treatment of chronic pancreatitis pain may involve sphincterotomy, pancreatic duct stenting, stone extraction, and drainage of a pancreatic pseudocyst. Therapy directed to the

pancreatic duct would seem to be most appropriate in the setting of a dominant stricture, especially if there is an obstructing intraductal stone. The use of endoscopic stenting for patients with chronic pain, but without a dominant stricture, has not been subjected to controlled trials. It is now appreciated that significant complications can occur from stenting (e.g., stent migration, stent occlusion, and stent-induced pancreatic duct strictures). Recent guidelines recommend considering celiac plexus block for treatment of pain in chronic pancreatitis, but recommendations were conditional with very low quality of evidence. Celiac plexus block has not been rigorously studied for chronic pancreatitis and does not provide durable pain relief. It can provide relieve in some selected patients, but the a priori identification of those who will respond is difficult. In patients with pancreatic duct dilation, ductal decompression with *surgical therapy* has been the therapy of choice. Among such patients, 80% seem to obtain immediate relief; however, at the end of 3 years, one-half of the patients have recurrence of pain. Two randomized prospective trials comparing endoscopic to surgical therapy for chronic pancreatitis demonstrated that surgical therapy was superior to endoscopy at decreasing pain and improving quality of life in selected patients with dilated ducts and abdominal pain. This would suggest that chronic pancreatitis patients with dilated ducts and pain should be considered for surgical intervention. The role of preoperative stenting prior to surgery as a predictor of response has yet to be proven.

Total pancreatectomy with or without autologous islet cell transplantation has been used in highly selected patients with chronic pancreatitis and abdominal pain refractory to conventional therapy. However, some patients will continue to have pain postoperatively, illustrating the complex nature of pain in patients with chronic pancreatitis. Patients who benefit most from total pancreatectomy have a shorter duration of symptoms and lower pain medication requirements. The role of this procedure remains to be fully defined but may be an option in lieu of ductal decompression surgery or partial pancreatic resection in patients with intractable, painful, small-duct disease or hereditary pancreatitis and particularly as the standard surgical procedures tend to decrease islet cell yield.

HEREDITARY PANCREATITIS

Hereditary pancreatitis (*PRSS1*) is a rare form of pancreatitis with early age of onset that is typically associated with familial aggregation of cases. A genome-wide search using genetic linkage analysis identified the hereditary pancreatitis gene on chromosome 7. Mutations in ion codons 29 (exon 2) and 122 (exon 3) of the cationic trypsinogen gene (*PRSS1*) cause an autosomal dominant form of pancreatitis. The codon 122 mutations lead to a substitution of the corresponding arginine with another amino acid, usually histidine. This substitution, when it occurs, eliminates a fail-safe trypsin self-destruction site necessary to eliminate trypsin that is prematurely activated within the acinar cell. These patients have recurring episodes of acute pancreatitis. Patients frequently develop pancreatic calcification, diabetes mellitus, and steatorrhea; in addition, they have an increased incidence of pancreatic cancer with a cumulative incidence of ~10%. A previous natural history study of hereditary pancreatitis in >200 patients from France reported that abdominal pain started in childhood at age 10 years, steatorrhea developed at age 29 years, diabetes at age 38 years, and pancreatic cancer at age 55 years. Abdominal complaints in relatives of patients with hereditary pancreatitis should raise the question of pancreatic disease.

PANCREATIC ENDOCRINE TUMORS

Pancreatic endocrine tumors are discussed in Chap. 84.

OTHER CONDITIONS

ANNULAR PANCREAS

When the ventral pancreatic anlage fails to migrate correctly to make contact with the dorsal anlage, the result may be a ring of pancreatic tissue encircling the duodenum. Such an annular pancreas may cause intestinal obstruction in the neonate or the adult. Symptoms of

postprandial fullness, epigastric pain, nausea, and vomiting may be present for years before the diagnosis is entertained. The radiographic findings are symmetric dilation of the proximal duodenum with bulging of the recesses on either side of the annular band, effacement but not destruction of the duodenal mucosa, accentuation of the findings in the right anterior oblique position, and lack of change on repeated examinations. The differential diagnosis should include duodenal webs, tumors of the pancreas or duodenum, duodenal ulcer, regional enteritis, and adhesions. Patients with annular pancreas have an increased incidence of pancreatitis and peptic ulcer. Because of these and other potential complications, the treatment is surgical even if the condition has been present for years. Retrocolic duodenjejunostomy is the procedure of choice, although some surgeons advocate Billroth II gastrectomy, gastroenterostomy, and vagotomy.

■ PANCREAS DIVISUM

Pancreas divisum is present in 7–10% of the population and occurs when the embryologic ventral and dorsal pancreatic anlagen fail to fuse, so that pancreatic drainage is accomplished mainly through the accessory minor papilla. Pancreas divisum is the most common congenital anatomic variant of the human pancreas. Current evidence indicates that this anomaly does not predispose to the development of pancreatitis in the majority of patients who harbor it. However, the combination of pancreas divisum and a small accessory orifice could result in dorsal duct obstruction. The challenge is to identify this subset of patients with dorsal duct pathology. Cannulation of the dorsal duct by ERCP is technically challenging and associated with a very high risk of post-ERCP pancreatitis, so patients with pancreatitis and pancreas divisum should likely be treated with conservative measures. In many of these patients, pancreatitis is idiopathic and unrelated to the pancreas divisum. Endoscopic or surgical intervention is indicated only if pancreatitis recurs and no other cause can be found. It should be stressed that the ERCP/MRCP appearance of pancreas divisum (i.e., a small-caliber ventral duct with an arborizing pattern) may be mistaken as representing an obstructed main pancreatic duct secondary to a mass lesion.

■ MACROAMYLASEMIA

In macroamylasemia, amylase circulates in the blood in a polymer form too large to be easily excreted by the kidney. Patients with this condition demonstrate an elevated serum amylase value and a low urinary amylase value. The presence of macroamylase can be documented by chromatography of the serum. The prevalence of macroamylasemia is 1.5% of the nonalcoholic general adult hospital population. Usually, macroamylasemia is an incidental finding and is not related to disease of the pancreas or other organs. Macrolipasemia has now been documented in patients with cirrhosis or non-Hodgkin's lymphoma. In these patients, the pancreas appeared normal on ultrasound and CT examination. Lipase was shown to be complexed with immunoglobulin A. Thus, the possibility of both macroamylasemia and macrolipasemia should be considered in patients with elevated blood levels of these enzymes.

A

This chapter represents a revised version of chapters by Drs. Norton J Greenberger (deceased), Phillip P. Toskes (deceased), Peter A. Banks, and Bechien Wu that were in previous editions of Harrison's.

■ FURTHER READING

- C SD et al: American Gastroenterological Association Institute guideline on initial management of acute pancreatitis. *Gastroenterology* 154:1096, 2018.
- F CE et al: Acute pancreatitis. *N Engl J Med* 375:1972, 2016.
- G TB et al: ACG clinical guideline: Chronic pancreatitis. *Am J Gastroenterol* 115:322, 2020.
- H PA, C DL: Chronic pancreatitis: Managing a difficult disease. *Am J Gastroenterol* 115:49, 2020.
- H PA et al: Recent advances in autoimmune pancreatitis. *Gastroenterology* 149:39, 2015.
- P MS, Y D: Global epidemiology and holistic prevention of pancreatitis. *Nat Rev Gastroenterol Hepatol* 16:175, 2019.
- Y D, L AB: The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology* 144:1252, 2013.